RTO-EN-2

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
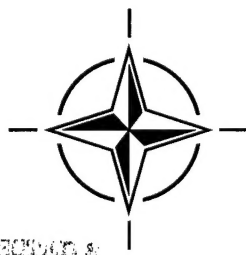
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

RTO LECTURE SERIES 214

Advanced Pattern Recognition Techniques
(Techniques avancées de reconnaissance de forme)

The material in this publication was assembled to support a Lecture Series under the sponsorship of the Sensors and Electronics Technology Panel and the Consultant and Exchange Programme of RTO presented on 14-15 September 1998 in Bristol, UK, on 17-18 September 1998 in Rome, Italy, and on 21-22 September 1998 in Lisbon, Portugal.

19981201 053

DTIC QUALITY INSPECTED 4

Published September 1998

Distribution and Availability on Back Cover

North Atlantic Treaty Organization

Research and **T**echnology **A**gency

RTA Headquarters: 7, rue Ancelle - 92200 Neuilly-sur-Seine, France

ST/60/4   19 August, 1998

TO: Recipients of RTO Publications

FROM: Scientific Publications Executive

SUBJECT: **RTO Technical Publications**

As you probably know, NATO formed the Research and Technology Organization (RTO) on 1 January 1998, by merging the former AGARD (Advisory Group for Aerospace Research and Development) and DRG (Defence Research Group). There is a brief description of RTO on page ii of this publication.

This new organization will continue to publish high-class technical reports, as did the constituent bodies. There will be five series of publications:

AG   **AGARDographs** (**A**dvanced **G**uidance for **A**lliance **R**esearch and **D**evelopment), a successor to the former AGARD AGARDograph series of monographs, and containing material of the same long-lasting value.

MP   **Meeting Proceedings**: the papers presented at non-educational meetings at which the attendance is not limited to members of RTO bodies. This will include symposia, specialists' meetings and workshops. Some of these publications will include a Technical Evaluation Report of the meeting and edited transcripts of any discussions following the presentations.

EN   **Educational Notes**: the papers presented at lecture series or courses.

TR   **Technical Reports**: other technical publications given a full distribution throughout the NATO nations (within any limitations due to their classification).

TM   **Technical Memoranda**: other technical publications not given a full distribution, for example because they are of ephemeral value only or because the results of the study that produced them may be released only to the nations that participated in it.

The first series (AG) will continue numbering from the AGARD series of the same name, although the publications will now relate to all aspects of defence research and technology and not only aerospace as formerly. The other series will start numbering at 1, although (as in the past) the numbers may not appear consecutively because they are generally allocated about a year before the publication is expected.

All publications, like this one, will also have an 'AC/323' number printed on the cover. This is mainly for use by the NATO authorities.

Please write to me (do not telephone) if you want any further information.

G.W.Hart

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

RTO LECTURE SERIES 214

Advanced Pattern Recognition Techniques
(Techniques avancées de reconnaissance de forme)

The material in this publication was assembled to support a Lecture Series under the sponsorship of the Sensors and Electronics Technology Panel and the Consultant and Exchange Programme of RTO presented on 14-15 September 1998 in Bristol, UK, on 17-18 September 1998 in Rome, Italy, and on 21-22 September 1998 in Lisbon, Portugal.

AQF99-02-0281

The Research and Technology Organization (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective coordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also coordinates RTO's cooperation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of initial cooperation.

The total spectrum of R&T activities is covered by 6 Panels, dealing with:

- SAS Studies, Analysis and Simulation
- SCI Systems Concepts and Integration
- SET Sensors and Electronics Technology
- IST Information Systems Technology
- AVT Applied Vehicle Technology
- HFM Human Factors and Medicine

These Panels are made up of national representatives as well as generally recognised 'world class' scientists. The Panels also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier cooperation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced
directly from material supplied by RTO or the authors.

Printed on recycled paper

Published September 1998

Advanced Pattern Recognition Techniques
(RTO EN-2)

Executive Summary

Pattern recognition has its roots in biological evolution. It is the extraction of consistent information from noisy spatiotemporal data and is currently being used in systems for battlefield supervision, smart weapons, anti-counterfeiting of all kinds, and for the reduction of false-alarm rates in the detection of land mines and unexploded ordnance.

Conventional methods of pattern recognition are mainly based on statistical approaches, such as density estimates or discriminant analysis. In this sense artificial neural networks can be regarded as an extension of these techniques. Fuzzy methods originate from control theory, but have also proven successful in pattern recognition. Over time, neuro-fuzzy methods have emerged that try to combine the advantages of each method and minimize the drawbacks.

An important task in pattern recognition is to choose the right features. Therefore a main part of this Lecture Series was devoted to feature extraction. This can be achieved in several ways: by electromagnetic and acoustic singularity expansion methods, by model based scattering signatures, also by using multiresolution time and frequency domain analysis, by SAR/ISAR imaging, bistatic microwave imaging and electromagnetic inversion techniques.

Practical applications of pattern recognition techniques were demonstrated with focus on statistical methods and artificial neural networks. Real-time software for discriminant and principal component analysis as well as PC based accelerator boards with on chip artificial neurons were introduced. Different methods for feature extraction with examples for automatic pattern recognition were shown.

The material in this publication was assembled to support a Lecture Series under the sponsorship of the Sensors and Electronics Technology Panel and the Consultant and Exchange Programme of RTO presented on 14-15 September 1998 in Bristol, UK, on 17-18 September 1998 in Rome, Italy, and on 21-22 September 1998 in Lisbon, Portugal.

Techniques avancées de reconnaissance de forme
(RTO-EN-2)

Synthèse

La reconnaissance de forme tire son origine de l'évolution biologique et peut être définie comme l'extraction d'informations cohérentes à partir de données spatio-temporelles brutes. Elle est utilisée pour la surveillance du champ de bataille, dans les munitions intelligentes, pour la contre contrefaçon et pour la rédution des taux de fausses alarmes dans la détection des mines terrestres et des munitions explosives non explosées.

Les méthodes classiques de reconnaissance de forme s'inspirent essentiellement d'approches statistiques, comme les estimations de densité et l'analyse discriminante. De ce point de vue, les réseaux neuronaux peuvent être considérés comme l'extension de ces techniques. Les méthodes floues dérivent de la théorie de commande, mais elles ont été employées avec succès pour la reconnaissance de forme. Suite à ces développements, sont apparus des neuro-méthodes ayant pour ambition de combiner les avantages de chaque méthode tout en réduisant au minimum leurs désavantages.

Le choix des caractéristiques appropriées est l'une des tâches essentielles de la reconnaissance de forme. Par conséquent, l'une des sessions principales de ce cycle de conférences a été consacrée à l'extraction des caractéristiques. Un certain nombre de techniques ont été examinées à savoir :
- les méthodes SEM acoustiques et électromagnétiques
- la modélisation des signatures de diffusion
- l'analyse dans les domaines temporel et fréquentiel
- l'imagerie SAR/ISAR
- l'imagerie hyperfréquence bistatique et les méthodes inverses

Des applications de techniques de reconnaissance de forme ont été présentées, l'accent étant mis sur les méthodes statistiques et les réseaux neuronaux artificiels. Des logiciels de gestion temps réel pour l'analyse discriminante et pour l'analyse des principaux composants ont été démontrés, ainsi que des cartes accélératrices pour PC intégrant des neurones artificiels sur puce. Différentes méthodes d'extraction de caractéristiques ont été exposées avec des exemples relatifs à la reconnaissance de forme.

Les textes contenus dans cette publication ont servi de support au Cycle de conférences 214 présenté sous l'égide de la Commission des senseurs et technologies de l'électronique dans le cadre du programme des consultants et des échanges de la RTO du 14 au 15 septembre 1998 à Bristol, au Royaume-Uni, du 17 au 18 septembre 1998 à Rome en Italie, et du 21 au 22 septembre 1998 à Lisbon au Portugal.

Contents

List of Authors/Speakers

Lecture Series Director: Professor Dr. J. NITSCH
University of Magdeburg
Institute of Electrical Engg & Power Electronics
P.O. Box 4120
D-39016 Magdeburg
GERMANY

Authors/Lecturers

Prof. Dr. Eric K. WALTON
Ohio State University
Electrical Engineering Dept., Electro Science Lab.
1320 Kinnear Road
Columbus, Ohio 43212-1191
UNITED STATES

Dr. Carl E. BAUM
Phillips Laboratory (PL/WSR)
3550 Aberdeen Avenue, SE
Kirtland AFB, NM 87117-5776
UNITED STATES

Prof. Dr. -Ing. habil. Hendrik ROTHE
University of the Federal Armed Forces Hamburg
Holstenhofweg 85
D-22043 Hamburg
GERMANY

Dr. A. Peter. M. ZWAMBORN
TNO Physics and Electronics Laboratory
Division Telecommunication and Defense Electronics
Group 3-3
Oude Waalsdorperweg 63
P.O. Box 96864
2509 JG's The Hague
NETHERLANDS

Co-Authors

Mr. Jan RHEBERGEN
TNO Physics and Electronics Laboratory
P.O. Box 96864
2509 JG The Hague
NETHERLANDS

Mr. André KASPER
University of the Federal Armed
  Forces Hamburg
Holstenhofweg 85
D-22043 Hamburg
GERMANY

Professor Dr. Joe LOVETRI
Dept. of Electrical & Computer Engineering
The University of Western Ontario
London, Ontario
CANADA N6A 5B9

Dr. Kamal BELKEBIR
Faculty of Electrical Engineering
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
NETHERLANDS

Prof. Dr. Jean-Charles BOLOMEY
Laboratoire des Signaux et Systèmes
CNRS/SUPELEC
Plateau du Moulon
Gif sur Yvette Cedex
FRANCE

Mr. M. MAGG
IABG mbH
Einsteinstr. 20
D-85521 Ottobrunn
GERMANY

Mr. Thomas RINDER
University of the Federal Armed
  Forces Hamburg
Holstenhofweg 85
D-22043 Hamburg
GERMANY

Mr. Beibei ZHANG
Dept. of Electrical & Computer Engineering
The University of Western Ontario
London, Ontario
CANADA N6A 5B9

Dr. Amélie LITMAN
Faculty of Electrical Engineering
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
NETHERLANDS

Prof. Dr. Peter M van den BERG
Laboratory of Electromagnetic Research
Faculty of Electrical Engineering
Centre for Technical Geoscience
Delft University of Technology
P.O. Box 5031, 2600 GA Delft
NETHERLANDS

Mr. Arno VON DER FECHT
University of the Federal Armed
  Forces Hamburg
Holstenhofweg 85
D-22043 Hamburg
GERMANY

Dr. Sergey PRIMAK
Dept. of Electrical & Computer
  Engineering
The University of Western Ontario
London, Ontario
CANADA N6A 5B9

Prof. Dr. Anton G. TIJHUIS
Faculty of Electrical Engineering
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven
NETHERLANDS

Dr. Jean-Michel GEFFRIN
Laboratoire des Signaux et Systèmes
CNRS/SUPELEC
Plateau du Moulon
Gif sur Yvette Cedex
FRANCE

Mr. Richard F. BLOEMENKAMP
Laboratory of Electromagnetic Research
Faculty of Electrical Engineering
Centre for Technical Geoscience
Delft University of Technology
P.O. Box 5031, 2600 GA Delft
NETHERLANDS

Approaches to Pattern Recognition

Hendrik Rothe
e-mail: rothe@unibw-hamburg.de
Tel.: 040/6541–2723 – Fax: 040/6541–2743
Mess- und Informationstechnik, FB-MB, Universitaet der Bundeswehr
Holstenhofweg 85, D–22043 Hamburg

1 Introduction

Pattern Recognition[1] per se covers a wide range of activities from many areas of science, engineering and everyday life. It has a long and respectable history within engineering directed to military applications. However, the cost of hardware to acquire the necesssary data (images or sensor signals) restricted its broad application for many years. Nowadays, it is possible for almost everybody to design and test even powerful automated pattern recognition systems.

Therefore there is an increasing need to understand fundamentals of pattern recognition techniques.

How could pattern recogniton be defined? A good first approximation could be:

> *Given some examples of complex signals and their correct classification, make correct decisions automatically for a stream of future examples.*

The roots of pattern recognition can be found in biological evolution, since many of us humans can, e.g.

- spot changing weather

- identify thousands of species (flowers, plants, animals)

- recognize faces and voices

In science and technology emerged literally thousands of pattern recognition tasks, like:

- diagnosing diseases

- identify types of vehicles, planes

- identify fingerprints and DNA profiles

- recognize handwritten characters and human voice

- picking optimal moves in certain situations (chess, war)

- identify incoming missiles from sensor signals

- detecting land mines and unexploded ordnance

While humans are able to do many of these tasks quite well, the desire is to construct *machines* performing these tasks cheaper, better, faster and of course, automatically. *Pattern recognition* is the engineering discipline of building such machines.

Because humans can perform many pattern recognition tasks very well, there has been for many years an interchange of ideas between engineers in the pattern recognition area and psychologists and physiologists doing research on human and animal brains. In the late 1950s the result of this cooperation was the *perceptron*, in the mid 1980s *artificial neural networks* occured. Both approaches left their biological roots and were studied by exact mathematical techniques with respect to their engineering performance.

Human pattern recognition is mainly *learnt*. It is not possible to describe the rules used to recognize a certain voice. On the other hand, biologists can

give rules to discriminate, e.g. between African and Indian elephant.

Imagine to taste a large number of unlabeled tasses of tea. After some training we will be able to detect a grouping in the teas. But we will need a *teacher*, to tell us that our favourite group is, e.g. Darjeeling GFOP, first flush.

The discovery of new groupings in sets of data is called *unsupervised pattern recognition*. The more common method uses a *training set*, also called *recognition set* with predefined groupings of data. This is called *supervised pattern recognition*. After the system has learned, it classifies future data, the *prediction set* into one or more sets of learned classes.

Pattern recognition should not be confused with *machine learning* which was fathered by the artifcial intelligence and computer science communities. Machine learning labels its classes in general only with *true* or *false*, while pattern recognition uses much more classes. In general, it can be said that machine learning mimics human reasoning by logical or binary operations and background knowledge may be exploited.

Now we can define the pattern recognition task in a more formal way. We assume a given set of $\mathcal{K}$ well defined classes and a classifier $\mathcal{C}$. For each example of a class there is a certain number of *features*, or *feature variables*. These features are fed into the classifier $\mathcal{C}$. After a – hopefully – short time the classifier responds:

- this example is from class $\mathcal{K}_i$

- this example is from none of the classes

- this example is too complicated for mee

The second category contains all *outliers*, while the third reports all *rejects* or *doubts*. The primary assessment of $\mathcal{C}$ will be its performance in terms of per cents of correct classification. The other aspect is the power of explanation. Users need to have confidence in the system before it will be accepted. For example, no one really cares if an automatic Zip code reading machine rejects an odd letter. But when a civilian airliner is classified as enemy aircraft and shot down, or an area is classified mine free, and explosions happen, serious questions are raised. Therefore in many cases black $\mathcal{C}$ boxes are totally inacceptable.

Very often, the most important aspect in the design of a classifier is to choose the right features. If the wrong data are measured the task may become very difficult or even unachievable. In this sense, advances in pattern recognition techniques have come from better features instead of more complex classifiers.

2 Feature Extraction

As mentioned in the preceding section, feature extraction is very important for correct pattern recognition techniques. In principle, there are two main approaches, namely induction and deduction.

Induction is from the historic point of view the older method. It was used before the advent of modern science. The point is that certain hypotheses about the object under study lead to the extraction of information – that possibly – describes the object or process in a correct manner. Clearly, experience is of fundamental importance in this case. However, there is one basic principle called *Occam's Razor*. It was found by the English scientist *William of Occam* int middle of the 14th century. It says:

> *Non sunt multiplicanda entia praeter necessitatem.*

With respect to pattern recognition we could translate as follows:

> *Use as few feature variables as possible to provide consistent classification.*

Occams Razor is starting point of modern statistical inference theory, which provides the possibility to assess feature variables with respect to their specific discriminatory power or information content. Examples for such methods are Principal Component Analysis and Discriminant Analysis.

It should be mentioned here that Artificial Neural Networks, Fuzzy Sets and Neuro–Fuzzy–Methods are not capable of assessing feature variables, nor can they give information about the stability of the classification process, or the probability of correct decisions.

Deduction is a method which is typical for modern science and engineering. It works as follows:

- Make a physical model of the object or process under study

- Choose the applicable fundamental laws of nature

- Make a mathematical model

- Solve the equations of the mathematical model either analytically or numerically

- Find the variables describing the interesting properties of the system – these are the feature variables

Deduction is always more desirable than induction because it provides *insight*. However, since many processes can not be modeled today by deduction, peaceful co–existence between the two approaches will last for a long time.
In this paper, only feature variables derived by induction will be used.

2.1   ... from Images[2]

Features derived from images should be invariant with respect to *shift*, *rotation* and *scale*. While shift invariance can be obtained by using *Fourier* transform techniques, rotational invariance requires the transformation of the image in the polar plane. Finally, scale invariance can be achieved by the *Mellin* transform. Complete invariance implies the use of all three approaches.

However, by the use of moments one can achieve complete invariance without complicated and time consumimg numerical computation.

That in mind, we can now deal with some simple feature variables.

Area and Perimeter   are commonly used desriptors for regions in a plane.

Euler Number   This number is a topological descriptor for binary images. It is defined to be the number of connected components minus the number of holes inside the connected components.

Position, Orientation, Symmetry   These are useful descriptors of objects within images. Position refers to the location of the object in the plane. The object's *centroid* (center of mass) is used to specify it's position. Orientation refers to how an object is situated in the plane. The object's *moment of inertia* is used to determine it's angle of orientation. Symmetry is the ratio of the minimum moment of inertia to the maximum moment of inertia. For binary images the feature symmetry is a rough measure of how elongated an object is. In this sense, a circle has a symmetry of equal to 1, a straight line of equal to 0.

Moment invariants   are image statistics that are independent of rotation, translation (shift) and scale. Moment invariants are uniquely determined by an image and vice versa. These properties facilitate pattern recognition by a great extent.
The moments invariants can be derived from the definitions of moments, centralized moments, and normalized centralized moments. Let f be a continuous function defined over $\mathbf{R}^2$. The moment of order (p,q) of f is defined by

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy$$

where $p, q \in \{0, 1, 2, \dots\}$. It has been shown that if f is a piecewise continuous function with bounded support, then moments of all orders exist, and, additionally, m_{pq} is uniquely determined by f and vice versa. The central moments of f are defined by

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q dx dy$$

with

$$\bar{x} = \frac{m_{10}}{m_{00}}$$

and

$$\bar{y} = \frac{m_{01}}{m_{00}}$$

The point $(\bar{x}, \bar{y})$ is called the image centroid which is the same as the center of gravity for a rigid body in a force field.
The discrete counterpart of the centralized moment of order (p, q) is given by

$$\mu_{pq} = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q dx dy$$

The normalized central moment η_{pq} is defined by

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$$

where

$$\gamma = \frac{p+q}{2} + 1$$

The following seven momets invariants developed by *Hu* are in the continuous case independent of rotation shift and scale. In the discrete some aberrations may occur, so one has to be careful.

$$\Phi_1 = \eta_{20} + \eta_{02}$$
$$\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
$$\Phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})$$
$$\times \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})$$
$$\times \left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$
$$\Phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2\right]$$
$$+ 4\eta_{11}(\eta_{30}\eta_{12})(\eta_{21} + \eta_{03})$$
$$\Phi_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{21})$$
$$\times \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]$$
$$+ (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})$$
$$\times \left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

Very often also so called form factors are used to extract information from images. A very popular one is *compactness*:

$$compactness = \frac{Perimeter^2}{4\pi \times Area}$$

3 Assessment of Features by Principal Component Analysis

The design matrix of a Principal Component Analysis[6,7,8] (PCA) can easily be defined by considering Table 3. For example, the $h = 1, \ldots, n$ objects or sample members are subjects, compounds, plants, technical products. The $k = 1, \ldots, p$ variables are time–dependent biological

It is also possible to use *Hough* transform for the detection of lines and statistical estimators of texture, like *energy, entropy, texture correlation, inverse difference moments*, and also *inertia*. This is, however, beyond our scope.

2.2 ... from Sensor Data

If a sensor is conceived as a device which transforms a certain physical, chemical, or biological property in a time dependent electrical signal, then we have to deal with one–dimensional problems. Because of the many advantages, we use again moments, but only one–dimensional ones. Since sensor data are always digitized, we have discrete data only. Therefore we are now in the domain of such statistics, which deals with measurements of a certain variable at a certain time. Table 1 displays so called empirical statistical moments of a one–dimensional probability distribution.

For all our classification examples, we will use these features.

Furthermore, Table 2 shows actual data of degradation measurements for high performance optics. There are four damage effects due to environmental problems which degrade image quality. This is the test set for all pattern recognition approaches discussed in this paper.[3,4,5]

Figure 1 shows the backscatter–curves for four different error pattern. They are significantly different from each other. There are 24 observations – 1 per sample – and namely 6 per error pattern. Five observations of every group form the recognition set $\mathcal{R}$ which is used for the supervised learning of the classification algorithm employed. Observations 1, 7, 13 and 19 are used for testing the predictive power of the classification approaches under study. The ocurrence of the 4 error pattern is in general equally likely.

responses, linear free energy–related (LFER) parameters, quantum chemical indices, substituent constants, descriptors of the quality of a product. Formally, the design is equivalent to that of an *one-group design* with a *single set of variables*.

No.	Name	Formula	Remarks
1	$Q(x_1 \ldots x_N)$	$\left(\dfrac{x_{min}}{x_{max}}\right)$	x_i is the measured backscatter–intensity at the sample point i
2	$x_M(x_1 \ldots x_N)$	$\left(\dfrac{1}{N}\sum\limits_{i=1}^{N} x_i\right)$	Average of all N measured values
3	$ADev(x_1 \ldots x_N)$	$\left(\dfrac{1}{N}\sum\limits_{i=1}^{N} |x_i - x_M|\right)$	Average Deviation More robust estimator than variance
4	$\sigma(x_1 \ldots x_N)$	$\left(\sqrt{\dfrac{1}{N-1}\sum\limits_{i=1}^{N}(x_i - x_M)^2}\right)$	Standard Deviation Square root of variance
5	$Var(x_1 \ldots x_N)$	$\left(\dfrac{1}{N-1}\sum\limits_{i=1}^{N}(x_i - x_M)^2\right)$	Variance Measure for the "variability" of data
6	$Skew(x_1 \ldots x_N)$	$\left(\dfrac{1}{N}\sum\limits_{i=1}^{N}[\dfrac{x_i - x_M}{\sigma}]^3\right)$	Skewness Characterizes the shape of the underlying distribution
7	$Kurt(x_1 \ldots x_N)$	$\left(\dfrac{1}{N}\sum\limits_{i=1}^{N}[\dfrac{x_i - x_M}{\sigma}]^4 - 3\right)$	Kurtosis Measure for "peakedness" or "flatness" of a distribution

Table 1: Feature variables extracted from backscatter–measurement

3.1 Assumptions

Design Robustness.
Please, take care that for most designs the degree of freedom due to error,

$$n_e = n - p - 1$$

is equal to or larger than the degree of freedom due to hypothesis,

$$n_h = p$$

The variables are continuously distributed or discrete random variables. Whenever possible, do not employ a *mixture of* measured *variables* and qualitative ones, because the resulting categorization may induce a grouping effect. For example, if a variable includes only ones and zeroes, two classes of sizes N_1 (the number of ones) and N_2 (the number of zeroes) may lead to artificial clustering, *mixtures of distributions*, etc. A careful check of the results is then necessary to avoid wrong conclusions.

Model Robustness.
The analysis is scale sensitive. Outliers are expected to have improportionally heigh weight in influencing the orientation of objects in a multidimensional pattern space. This is an advantage because unusual effects can so be discovered very rapidly.

Hypothesis Testing.
The method is perfectly general; it involves no assumptions on the underlying distribution model. Inasfar, it is a nonparametric approach. However, for hypothesis testing, it is only optimal if multivariate normality is assumed. Inasfar, the *probabilistic* principal component analysis is parametric.

Effect	Group	Obs.	$\mathcal{P}/\mathcal{R}$	Q	Ave	$ADev$	$SDev$	$SVar$	$Skew$	$Kurt$
		1	$\mathcal{P}$	-20.32	-13.20	1.47	2.08	4.34	-0.26	4.87
		2	$\mathcal{R}$	-17.43	-10.83	0.79	1.37	1.87	3.39	26.88
Scating	1	3	$\mathcal{R}$	-16.85	-9.86	1.04	1.61	2.60	2.77	13.81
		4	$\mathcal{R}$	-16.42	-9.35	0.86	1.35	1.83	3.49	22.95
		5	$\mathcal{R}$	-16.85	-9.91	0.90	1.41	2.00	2.85	20.16
		6	$\mathcal{R}$	-16.94	-10.29	0.82	1.37	1.86	3.53	25.95
		7	$\mathcal{P}$	-14.60	-6.17	0.63	1.07	1.14	4.90	44.00
		8	$\mathcal{R}$	-14.66	-6.29	0.59	1.06	1.12	4.59	44.82
Fog	2	9	$\mathcal{R}$	-14.60	-6.12	0.62	1.09	1.18	4.70	40.89
		10	$\mathcal{R}$	-14.49	-5.73	0.54	0.97	0.94	4.99	50.98
		11	$\mathcal{R}$	-14.77	-6.19	0.61	1.05	1.10	3.79	39.56
		12	$\mathcal{R}$	-14.70	-6.02	0.61	1.06	1.12	3.84	37.33
		13	$\mathcal{P}$	-14.44	-6.30	1.25	1.57	2.46	1.73	7.98
		14	$\mathcal{R}$	-14.68	-6.66	1.31	1.62	2.61	1.64	7.11
Holes	3	15	$\mathcal{R}$	-14.31	-6.23	1.26	1.57	2.45	1.76	8.15
		16	$\mathcal{R}$	-14.22	-5.68	1.63	1.92	3.70	0.94	1.60
		17	$\mathcal{R}$	-14.26	-6.38	1.32	1.62	2.63	1.63	6.55
		18	$\mathcal{R}$	-14.32	-6.44	1.29	1.61	2.58	1.71	7.04
		19	$\mathcal{P}$	-14.50	-7.73	0.47	0.93	0.87	8.66	109.28
		20	$\mathcal{R}$	-15.63	-9.03	0.67	1.12	1.24	5.28	54.42
Scating	4	21	$\mathcal{R}$	-14.70	-8.05	0.46	0.93	0.87	8.93	112.31
+		22	$\mathcal{R}$	-15.25	-8.66	0.53	1.04	1.08	7.17	80.50
Fog		23	$\mathcal{R}$	-15.96	-9.58	0.57	1.09	1.18	6.64	72.31
		24	$\mathcal{R}$	-15.57	-8.78	0.67	1.19	1.43	4.97	44.10

Table 2: Classification data ($\mathcal{P}$: prediction set, $\mathcal{R}$: recognition set)

3.2 Goal of Analysis

1. The original data should be transformed to asymptotically normally distributed scores with unit variances and zero means in order to make variables with different variances and/or meaning commensurable, regardless of the fact that the original measurements may be "distorted" when all variables are "treated equally". Furthermore, the resulting *new* variables should be independently distributed (*completely* uncorrelated, orthogonal). This allows a better interpretation by illustrating graphically the results, and has strong advantages if subsequent analyses should be made.

2. The number of original variables should be reduced to satisfy the principle of parsinomy (to increase the test power of subsequent analyses) *without* remarkable loss of information. Thus, the *new* variables *should* be arranged in order of decreasing contributions to the information content of the variables to get a hierarchy ("dominant is first, least is last").

3. Internal relationships within the original variables, and between the original and new variables, should easily be discovered.

3.3 Application Examples

Application examples are so many–fold that only some may here be collected.

1. Discovery of an unknown, common, *synthetic* factor (also called metameter, composite parameter, overall parameter) of a profile of variables. Usually, the new parameter may be regarded as linear combination of the original variables. In such case, it is the first (dominant) principal component function. It should be emphasized that *linearity* concerns
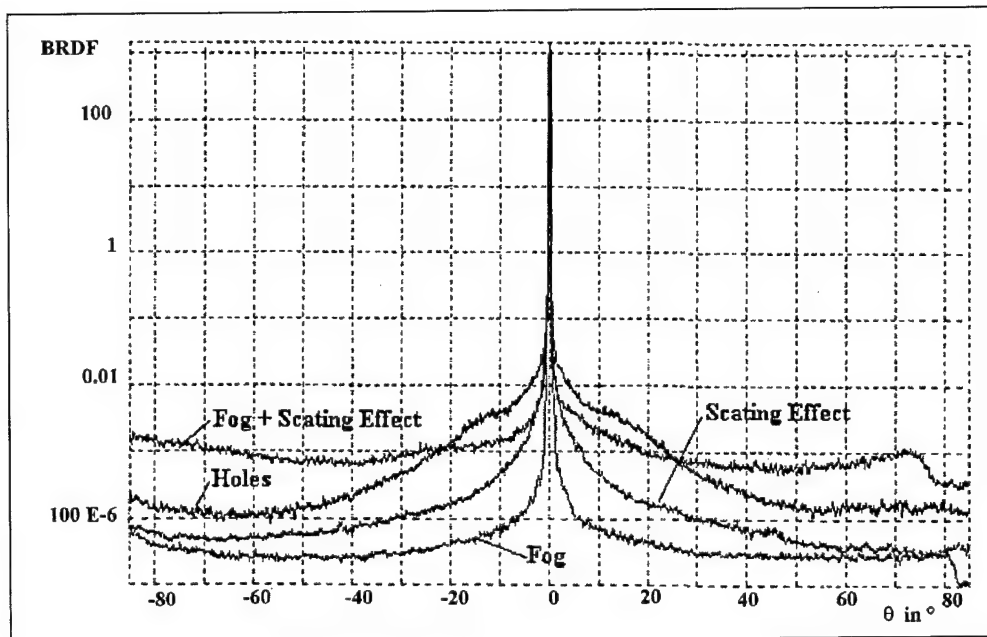
Figure 1: Typical backscatter–curves of the four kinds of surface damage

Objects	Y_1	Y_2	$\cdots$	Y_p
1	y_{11}	y_{12}	$\cdots$	y_{1p}
2	y_{21}	y_{22}	$\cdots$	y_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{n1}	y_{n2}	$\cdots$	y_{np}

Table 3: Design Matrix of PCA

only the mechanics.

For example, the variables y, y^2 enter in linear combination if the transformation $Y_1 = y, Y_2 = y^2$ is applied, yielding

$$\mathcal{F}(Y_1, Y_2) = a_0 + a_1 Y_1 + a_2 Y_2$$

where a_0 to a_2 are coefficients to be estimated.

2. Discovery of the structure of a single variable which may be thought as a mixture of *intercorrelated* variables.

3. Application of diagnostic statistics, such as of (i) outlying observations, (ii) multicollinear data structures, (iii) homogeneous subsets of objects (clusters, groups) that are scattered around a central value, (iv) homogeneous subsets of variables.

4. Use of the components in subsequent analyses, such as in (i) regression analysis and (ii) *Euclid*ean distance analysis (in this case, the distances are invariant under rotation of the axes).

3.4 Algorithms

3.4.1 Underlying Formalisms

1. Let $\mathbf{Y}$ be the (n, p)–matrix of the design of Table 3,

$$\mathbf{Y} = (y_h) = (y_{hk})$$

where $h = 1, \dots, n$ denotes the objects, $k = 1, \dots, p$ denotes the number of variables that determine the *dimensionality* of a distribution. The matrix may also be written in terms of the Y_k variables,

$$\mathbf{Y}^T = (Y_1 \dots Y_p)$$

where $\mathbf{Y}^T$ denotes the transposed matrix, and the parentheses symbolize an *order* of variables. In contrast, $\{\dots\}$ denotes an arbitrary sequence of parameters. The so-called *within–group matrix* becomes then in this one–sample design

$$\mathbf{W} = \sum_{h=1}^{n} (\mathbf{y}_h - \bar{\mathbf{y}})(\mathbf{y}_h - \bar{\mathbf{y}})^T$$

where the p–dimensional vector of arithmetic *mean values* is defined by

$$\bar{\mathbf{y}} = \frac{1}{n} \sum_{h=1}^{n} \mathbf{y}_h$$

The (p, p)–dimensional variance–covariance or *dispersion matrix* becomes then

$$\mathbf{S} = \frac{1}{n-1} \mathbf{W}$$

where the main–diagonal elements contain the variances, and the off–elements include the covariances. This matrix plays a fundamental role in all approaches. Its diagonal matrix may be written as

$$diag(\mathbf{S}) = \mathbf{D}$$

(it includes the variances in the main diagonal while the off–elements are zeroes). The *correlation matrix* $\mathbf{R}$ is the second matrix of fundamental importance; it is defined by

$$\mathbf{R} = (\mathbf{D}^{-\frac{1}{2}} \mathbf{S})(\mathbf{D}^{-\frac{1}{2}})$$

In this special case, $\mathbf{R}$ is a symmetric (p, p)–matrix which simplifies considerably the computation of its eigenvalues and eigenvectors.

As already described, the $f = 1, \dots, s \leq p$ non–zero *eigenvalues* λ_f are ordered, and its resulting s–dimensional vector may be written as

$$\lambda^T = (\lambda_1 \lambda_2 \dots \lambda_s)$$

Let $\mathbf{I}$ be a (p, p)–dimensional identity matrix (a matrix where the main diagonal contains ones while the off–elements include zeroes). Then, the eigenvalues are obtained by

$$\det(\mathbf{R} - \lambda \mathbf{I})\mathbf{v}_f = \mathbf{0}$$

where det denotes the determinant. We remember that the determinant of $\mathbf{R}$ is given by

$$\det(\mathbf{R}) = \prod_{f=1}^{p} \lambda_f$$

and the determinant of the inverse $\mathbf{R}^{-1}$ is

$$\det(\mathbf{R}^{-1}) = \prod_{f=1}^{p} \frac{1}{\lambda_f}$$

These properties are useful for procedures applied to other techniques than PCA. The *eigenvectors* $\mathbf{v}_f$ are obtained by

$$(\mathbf{R} - \lambda_f \mathbf{I}) = 0$$

where the normalization condition is

$$(\mathbf{v}_f)^T \mathbf{v}_f = 1$$

because $\mathbf{R}$ is a symmetric matrix. Remember also that the angle between two eigenvectors, say, for example, $\mathbf{v}_\alpha$ and $\mathbf{v}_\beta$ can easily be determined by

$$\cos(\mathbf{v}_\alpha, \mathbf{v}_\beta) = \frac{\mathbf{v}_\alpha^T \mathbf{v}_\beta}{\|\mathbf{v}_\alpha\| \, \|\mathbf{v}_\beta\|}$$

For example, let $\mathbf{v}_\alpha$ be the *first* eigenvector of a principal component analysis of group $\mathcal{A}$, and let $\mathbf{v}_\beta$ be the *first* eigenvector of group $\mathcal{B}$. This offers the *possibility to compare groups* by angular statistics, but that is not a matter of concern here.

2. The second step is the calculation of the matrix of *coefficients of principal component functions*,
$$\mathbf{P}^T = \mathbf{L}^{-\frac{1}{2}} \mathbf{V}$$

where $\mathbf{L}$ is the diagonal matrix of the eigenvalues, and $\mathbf{V}$ is the matrix of eigenvectors. As already mentioned, there are good reasons to make the original measurements commensurable. This may be done by the transformation

$$\hat{y}_h = \mathbf{D}^{-\frac{1}{2}} (\mathbf{y}_h - \bar{\mathbf{y}}).$$

The resulting matrix $\hat{\mathbf{Y}}$ has unit variances and zero means. The *scores* of the principal component function are then computed by

$$\mathbf{F} = \bar{\mathbf{Y}} \mathbf{P}$$

and the covariances resp. correlations between the components $\mathbf{F}_f$ are zero. In other words, the dispersion matrix of the *components* is an identity matrix.

Spectral decomposition of the correlation matrix means that $\mathbf{R}$ can be expressed by

$$\mathbf{R} = \sum_{f=1}^{s} \lambda_f \mathbf{v}_f \mathbf{v}_f^T$$
$$= \mathbf{V} \mathbf{L} \mathbf{V}^T$$

If only the "relevant" eigenvalues and eigenvectors are used, the resulting correlation matrix $\mathbf{R}_{\text{deg}}$ deviates slightly from $\mathbf{R}$, and the difference

$$\mathbf{R}_e = \mathbf{R} - \mathbf{R}_{\text{deg}}$$

is the *residual matrix* which should only contain zero or near–zero values.

3. In a third stage of numerical computation, the *correlation matrix* $\mathbf{M}$ *between original variables and components* may be estimated directly or by the simple equation

$$\mathbf{M}^T = \mathbf{L}^{\frac{1}{2}} \mathbf{V}$$

Re–arranging the equations, we can now return to the original data matrix if all components are included, or to a degenerated or *adjusted original matrix* if at least one component is omitted,

$$\mathbf{Y} = (y_{hk}) = \bar{y}_k + s_k(q_{hk})$$

$$(k = 1, \ldots, p)$$

where $\bar{y}_k$ and s_k are the kth mean value and standard deviation (square root of the kth variance) of the original measurements, and

$$\mathbf{Q} = (q_{hk}) = \mathbf{F} \mathbf{M}^T$$

The adjusted variables may loosely be compared with the adjusted means of covariance analysis.

4. The squared distance between two objects, termed $\mathbb{A}$ and $\mathbb{B}$ for simplicity, is the squared *Euclidean distance*

$$d^2(\mathbb{A}, \mathbb{B}) = \|\mathbf{F}_{\mathbb{A}} - \mathbf{F}_{\mathbb{B}}\|^2$$

Thus, principal component analysis may also be applied for discovering subsets of objects around the gravity center or other centers, and has strong relationships to *cluster analysis* and *multidimensional scaling* (principal coordinate analysis). It should be emphasized that the number of components applied to an actual analysis can be determined by the user himself.

3.4.2 Decision Procedure

Test of Dimensionality of Original Variables
The squared multiple correlation coefficients *within* the variables, called *internal determination coefficients*,

$$D_k = 1 - \Lambda_k = 1 - 1/r^{kk}$$

$$(k = 1, \ldots, p; 0 \le D_k \le 1)$$

must be significantly different from zero if subsequent analysis makes sense. Note that r^{kk} is the kth element of the main diagonal of $\mathbf{R}^{-1}$, and Λ_k is the likelihood–ratio criterion.

It can be shown that, if at least one eigenvalue λ_f of $\mathbf{R}$ is zero, at least a single D_k is one. The null hypothesis $\mathbb{H}_0 ::=$ (at least one internal determination coefficient of the *population* is zero) is rejected at the significance level α if the test statistic TS is

$$TS = D_k \ge (1 - \Lambda_\alpha) = \theta_\alpha$$

where Λ_α is the likelihood–ratio criterion, θ_α is the largest–root criterion, and the degrees of freedom are here

$$n_h = p - 1, n_e = n - p$$

Rejecting this null hypothesis suggests that *at least* one multiple correlation coefficient of the original data is significant.

Test of Dimensionality of Components The big question of PCA is how the essential dimensionality should be determined. We have solved this problem by testing the eigenvalues of $\mathbf{R}$ and the correlation coefficients of $\mathbf{M}$.

a. *Test of Eigenvalues.* Let λ_f be the estimator of the eigenvalue σ_f of the population. Then, the *global null hypothesis* becomes

$$\mathbb{H}_{01} ::= (\sigma_1 = \ldots = \sigma_s).$$

This is equivalent to the assumption that the correlation matrix of the population is an identity matrix. "Accepting" this null hypothesis would suggest the existence of a *spherical symmetry* or, respectively, rejecting this hypothesis would suggest that *at least* one correlation coefficient *of the original data* is significant.

The null hypothesis is rejected at the significance level α if

$$TS = q \ln(\alpha_0/\beta_0)^r \ge \chi^2_{\mu;\alpha}$$

where (α_0/β_0) is the likelihood–ratio criterion. The test statistic is asymptotically χ^2 with

$$\mu = \frac{1}{2}(p - K + 2)(p - K - 1)$$

degrees of freedom, and K is the number of eigenvalues to be removed ($K = 0$ for testing $\mathbb{H}_{01}$). Furthermore, we obtain

$$\alpha_0 = \frac{1}{(p - K)} \sum_{k=0}^{p} \lambda_K$$

and

$$\beta_0 = \left(\prod_{K=0}^{p} \lambda_K\right)^{1/(p-K)}$$

Thus, for testing $\mathbb{H}_{01}$, the procedure is simplified to $\alpha_0 = 1$ and

$$\beta_0 = \det(\mathbf{R})^{\frac{1}{(p-K)}}$$

We must still define q and r. For *small sample sizes*, the two arguments are defined as follows:

$$r = 1, q = (n - 1)(p - K)$$

For larger samples, a more conservative decision may be applied:

$$r = p - K, \qquad q = (n-1) - \frac{1}{2}(2p+5) - \frac{2}{3}K$$

If $\mathbb{H}_{01}$ is rejected, we can test the significance of the individual eigenvalues resp. the *local null hypotheses*,

$$\mathbb{H}_{02} ::= (\sigma_1 = 0, \sigma_2 = 0, \ldots, \sigma_s = 0)$$

First, λ_1 is excluded ($K = 1$), and TS and μ must be determined as described above. The desired test statistic for examining $\mathbb{H}_{02} ::= (\sigma_1 = 0)$ is then given by the difference of TS related to $K = 0$, and of TS related to $K = 1$. In strict analogy, the difference of the degrees of freedom is determined. Then, the χ^2 significance point is taken from tables, and the procedure is continued if the first

eigenvalue is significant. In strict analogy, the subsequent eigenvalues are examined.

Remark. Remember the test of significance of eigenvalues is sensitive to departures from normality.

The information content of an individual component can be visualized by

$$\inf(\%)Pca = \lambda_f(100\%)/tr(\mathbf{R})$$

Because $\mathbf{R}$ is a correlation matrix, its trace is equivalent with the sum of eigenvalues or the number s of non–zero eigenvalues,

$$tr(\mathbf{R}) = \sum_{f=1}^{s} \lambda_f = s$$

and the cumulative percentage cum(%) PCA is its successive sum.

b. *Test of Correlation Coefficients.* Up to now, a joint probability density function of sample correlation coefficients is not available in a closed form. However, *a–priori* statistics can be employed.

Setting ρ for any of the coefficients of a correlation matrix, and let r be its estimator. Then, the null hypothesis

$$\mathbb{H}_{03} ::= (\rho = 0)$$

is rejected at the significance level α if the test statistic is

$$TS = |\, r \,| \geq g_0 = \left(t^2_{f_1, f_2; \alpha}/(t^2_{f_1, f_2; \alpha} + f_2) \right)^{\frac{1}{2}}$$

where the degrees of freedom are

$$f_1 = p, f_2 = n - p - 1$$

with p as number of variables under study.

c. *Diagnostic Statistics*

1. **Normality**. In general, it can be stated that parametric methods are relatively robust against departures from normality if the principles of design robustness are satisfied approximately. Nevertheless, in some cases it is necessary to examine statistically this assumption. A rejection of the hypothesis of normality means:

(i) there is a non-normality *or*

(ii) we have a mixed distribution consisting of unimodal distributions with different variances resp. dispersion matrices (grouped observations *within* a one–sample design), *or*

(iii) points (i) *and* (ii) are valid. In particular, we find this property if *one–group designs* contain various *subgroups* (clusters with homogeneous distributions).

For a *small sample size* ($n \leq 30$), the scores are ranked and then regressed against the ordinate values of the cumulative normal distribution (95% confidence limit). If the correlation coefficient is significant, the intercept is approximately 50%, and a common slope can be observed, normality may be hypothesized.

For *larger sample sizes*, a frequency histogram shows a bell-shaped distribution, that is, the cumulative frequencies *vs.* the upper class limits leads to a straight line in the normal probability paper of *Hazen*. Rules are available for the assignment of sample members into classes and the estimation of class limits. Another possibility is the estimation of skewness g_3 and kurtosis g_4. For large samples (asymptotic case), the skewness becomes:

$$skew. = \frac{n \sum_{h=1}^{n}(y_h - \bar{y})^3}{(n-1)(n-2)s^3}$$

(missing if the standard deviation s is zero, or of the sample size n is less than 3). We use the standardized skewness,

$$g_3 = \frac{skew.}{\sqrt{\frac{6}{n}}}$$

because it is asymptotically normally distributed (for example, an *absolute* value of g_3 that is equal to or larger than

1.96 means that the hypothesis of normality must be rejected at the 5% significance level or less). For large samples, the kurtosis becomes

$$kurt. = \frac{n(n+1)\sum_{h=1}^{n}(y_h - \bar{y}^4)}{(n-1)(n-2)(n-3)s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

(missing if $s = 0$ or $n < 4$). We propose to use the standardized version,

$$g_4 = \frac{kurt.}{\sqrt{\frac{24}{n}}}$$

because this statistic is asymptotically normally distributed.

We employ also a suitable compromise between the two procedures. It is supposed that y_{hk} represents $h = 1, \ldots, n$ independent observation of a continuous random variable Y_k for each of the $k = 1, \ldots, n$ variables, with a cumulative distribution function (*cdf*)

$$\mathcal{F}\left((Y_k - \mu_k)/\sigma_k\right)$$

where μ_k and σ_k are the population means and standard deviations. The probability plot is a scatter of the ordered sample values

$$y_{1k} \leq \cdots \leq y_{nk} \text{vs.} z_{1k} \leq \cdots \leq z_{nk}$$

where

$$z_{hk} = \mathcal{F}^{-1}(\pi_{hk})$$

with π_{hk} as an empirical estimate of the *cdf*. Several values for π_{hk} have been suggested; we use here

$$\pi_{hk} = \frac{(h - \frac{3}{8})}{(n + \frac{1}{4})}$$

(for each $k = 1, \ldots, p$)

If $\mathcal{F}$ is the *cdf* of a normal distribution, the resulting scatter plot is called *normal probability plot*, and the z_{hk}

are called normal scores. Statistically speaking, normality must be hypothesized if the correlation coefficient between the ranked original data and the normal scores is significant and if there is a common straight line. It must be emphasized that a *graphical plot* is an important means to get a survey on the role of individual sample members (outliers, clustering etc.).

Deviations from normality have usually no serious consequences if design and model robustness exists, and if there is a symmetry of distribution. In some cases, an apparent normality may consist of a *mixture of distributions*.

Compensating for non–normality is to employ a transformation (which normalizes and works often as variance–stabilizing transformation, too) to each variable. Although the effect to *multivariate normality* is general unknwon and this univariate transformation may only achieve marginal normality, it has the advantage of simplicity. The limitation of tests of *multivariate normality* is that their power is still to be shown; furthermore, tables of critical quantiles include only the uni– and bivariate case, in general. Thus, it is difficult to test multivariate normality with the exception of very particular designs. We propose to examine a jointly distributed composite parameter, such as scores of the significant *principal component functions*.

2. *Randomness of Data*. Randomness of data is the basis of considering each selection of objects to be equally likely. Various tests are available. We employ the serial correlation coefficient. In most cases, randomness is obtained by experimental designing so that mutual dependence of the sample members (autocorrelation) can be excluded. In some cases, a measurement or residual of regression analysis may depend on the previous values. *Examples*: Growth curves, time series, nonlinear models,

step–by–step optimization procedures (the result depends then on the previous results). The estimators will no longer be minimum variance estimators. *Compensating* for randomness is based on the use of random number tables, and of autoregressive transformations.

3. *Outliers.* Let $\mathbf{R}_h$ be any correlation matrix *without* the sample member $-h$, let $\mathbf{R}$ be the correlation matrix of all sample members. Then, the smallest scattering ratio

$$R_{-h} = \min\left\{ \frac{\det(\mathbf{R}_{-h})}{\det(\mathbf{R})} \right\}$$

is an indicator that $-h$ is an outlier (the range of R_{-h} varies from 0 to 1). If the empirical value R_{-h} is *less than* the tabulated quantile r_α, the null hypothesis (no outlier) must be rejected. It should be emphasized that this test is very radical ("very sharp").

3.5 Computational Examples

3.5.1 Test example

The question is now, how PCA can be used as a tool for machine learning and pattern recognition. The way we employed is pointed out by the following small example. Imagine that we have four different pattern in our data set. These pattern can be conceived to be obtained from the same input data X which have been subjected to certain *data transformations* $\mathcal{T}$. For simplicity we choose:

- $\mathcal{T}_\alpha \equiv Y$ (i.e. the output data are equal to the input data)

- $\mathcal{T}_\beta \equiv Y^2$

- $\mathcal{T}_\gamma \equiv \sqrt{Y}$

- $\mathcal{T}_\delta \equiv \ln(Y)$

Using the set of input data $Y \sim (1, 2, 3)$ the (4,3)–design matrix of PCA becomes:

$$\mathbf{Y} = \begin{pmatrix} 1.00 & 2.00 & 3.00 \\ 1.00 & 4.00 & 9.00 \\ 1.00 & 1.41 & 1.73 \\ 0.00 & 0.69 & 1.09 \end{pmatrix}$$

Figure 2 shows a listing in extracts produced by the software package MDAD: *Multivariate Design, Analysis and Diagnosing* which was developed in the research group of the author. The signs of the 1st and 2nd principal component are:

$$\mathbf{C} = \begin{pmatrix} + & - \\ + & + \\ - & - \\ - & + \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{T}_\alpha \\ \mathcal{T}_\beta \\ \mathcal{T}_\gamma \\ \mathcal{T}_\delta \end{pmatrix}$$

where $\mathbf{C}$ is the classification matrix. The four different pattern are fully recognized. The features are consistent in this simple case.

```
MDAD  1.02a (c) --- Multivariate Statistics
         Principal Component Analysis
```

Matrix TEST-EXPL.MAT

	1	2	3
1	1.00	2.00	3.00
2	1.00	4.00	9.00
3	1.00	1.41	1.73
4	0.00	0.69	1.10

Means

0.75	2.03	3.71

Standard Deviations

0.50	1.42	3.62

Correlation Matrix(c0 = 1.00)

```
                  1              2              3
      1    1.00           0.63           0.48
      2    0.63           1.00           0.98
      3    0.48           0.98           1.00


  Eigenvalues
           2.42  *        0.58           0.96e-03


  Eigenvectors
                1              2              3
      1    0.48           -0.86          0.15
      2    0.63            0.23          -0.74
      3    0.60            0.45          0.66


  Principal Component Function
                1              2              3
      1    0.31           -1.13          4.75
      2    0.41            0.30          -23.89
      3    0.39            0.59          21.24


  Principal Component Function Scores
                1              2              3
      1    0.07           -0.69          -1.33
      2    1.29            0.72          0.26
      3   -0.23           -1.02          1.08
      4   -1.13            0.99          -6.82e-03


The global Null Hypothesis is rejected: TS (19.83) > ChiSquare (15.09)
General Alpha level is equal to 1% (Values marked by an asterisk are
significant)
Test of Eigenvalues was performed automatically
Number of Non-Zero Eigenvalues is      3
Number of significant Eigenvalues is 1
```

Figure 2: Listing in extracts

3.5.2 Example

Figure 3 shows classification results for the example according Table 2.
The statistical moments were computed according to Table 1, and the resulting design matrix was subjected to a principal component analysis. The signs of principal components were used for classification, no error occured. Furthermore the results of PCA indicate that the features form a consistent design matrix from the *statistical* point of view. That means, *all pattern recognition approaches should be able to detect the four error pattern.*

4 Pattern Recognition Methods

4.1 Discriminant Analysis

4.1.1 General purpose and computational approach

General purpose Discriminant analysis[9,10] is commonly used to determine which variables discriminate best between two or more groups. The basic idea underlying discriminant analysis is to determine whether groups differ with regard to the mean of a feature variable. This variable is then used to predict group membership.

Computational approach Computationally, discriminant analysis is very similar to the analysis of variance (*ANOVA/MANOVA*). In the case of a single variable, the final significance test of whether or
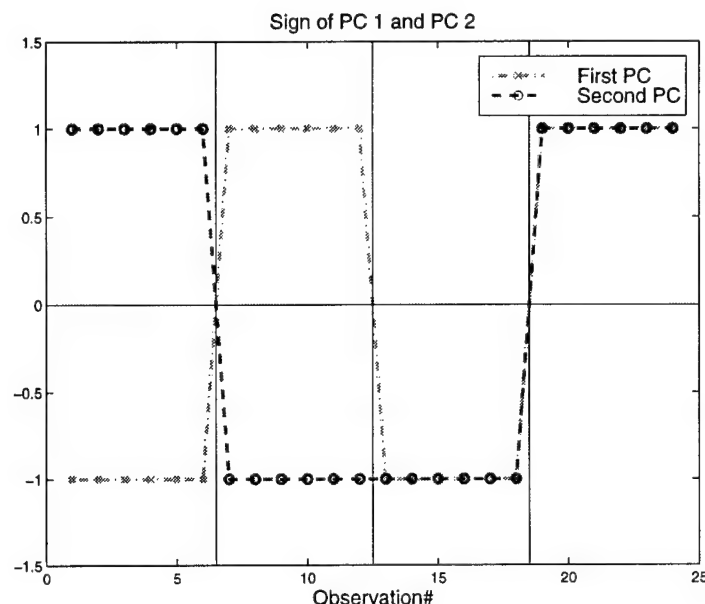
Figure 3: 1st and 2nd principal component vs. No. of observation

not a variable discriminates between groups is the F–test. F is computed as the ratio of the between–groups variance in the data over the pooled within–group variance. If the between–group variance is significantly larger than the within–group variance then there must be significant differences between means. Usually, several variables are included in a discriminant analysis – we actually use 7 feature variables in our example. Consequently, we have a matrix of total variances and covariances, and also a matrix of pooled within–group variances and co-variances. These two matrices can be compared via multivariate F–tests in order to decide whether or not there are any significant differences between groups. Even though the computations with multiple variables are more complex, the principal reasoning is still the same, and namely that we are looking for significant mean differences between groups.

Before starting the computational procedures it makes always sense to have a look at the data profiles. This gives a first impression, if discrimination of multiple groups could be possible. In Figure 4 the 7 feature variables are displayed in form of "data canyons".

It can be seen that the effects "Scating" and "Holes" have rather different profiles, while the separation of "Fog" and "Scating + Fog" could

be difficult. The main difference which is clearly recognizable is the form of the "ground" of the "canyon". In the case of "Scating + Fog" the ground is shaped like a gable, while the "Fog" profiles show a straight line tilted from left to right. However, discrimination of all 4 groups should be possible.

Canonical analysis When actually performing a multiple group discriminant analysis the use of canonical discriminant analysis will provide some optimal combinations of groups and variables with respect to best discrimination between groups. In this sense, the first discriminant function provides the most overall discrimination between groups, the second provides second most, and so on. Moreover, the discriminant functions will be *orthogonal*, i.e. their contributions to the discrimination of groups will not overlap.

4.1.2 Assumptions

Normality It is assumed that the measured data (backscatter–curves) and the derived feature variables represent a sample from a multivariate normal distribution. However, violations of the normality assumption are usually not severe, meaning that the results of analysis are still reliable.

Homogeneity It is assumed that the variance/covariance matrices are homogenous across groups. The multivariate *Box* M–test can be used. If deviations from homogeneity are detected, this should not be taken too seriously. However, our data are homogenous.

Correlations between means and variances Validity of discriminant analysis can be influenced when means and variances across groups are correlated. If there is a large variability in a group with particularly high means on some variables, then those means are not reliable. Significance tests will fail in this case. The experimental data of our example do *not* show this property.

Categorized normal probability plot This plot is constructed as follows. First, the deviations from the mean, i.e. the residuals, are rank ordered. From these ranks standardized values of the normal distribution based on the assumption that the data come from a normal distribution are computed. These values are plotted on the y–axis. *If* the observed residuals which are plotted on the x–axis are normally distributed then they will *not* deviate from the line. If there is a general lack of fit, and the data do not form a clear pattern, then non–normality may be assumed.
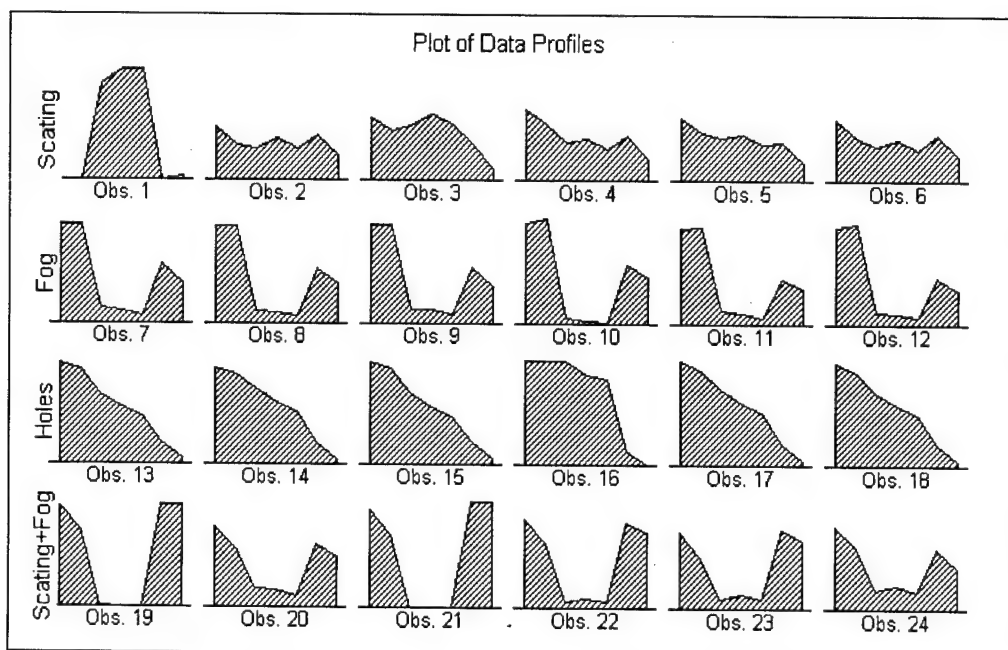
Figure 4: Data profiles

$\mathcal{UCDF}$ No.	Perc. of Var.	Coefficients for						
		Intercept	Q 1	Ave 2	$ADev$ 3	$SDev$ 4	$SVar$ 5	$Skew$ 6	$Kurt$ 7
1	60.79	-67.02	2.22	-2.43	45.61	72.55	-32.03	-0.48	0.20
2	30.96	-94.98	-6.28	6.73	1.63	71.59	-24.21	-1.75	0.18
3	8.25	-195.26	-11.69	6.57	7.62	75.24	-24.98	0.86	0.11

Table 4: Unstandardized canonical discriminant function coefficients

Group	Value of $\mathcal{UCDF}$ No.		
No.	1	2	3
1	1.1860	-4.1838	4.1985
2	-10.4797	5.3353	0.0329
3	10.4154	5.1644	-0.8131
4	-1.1217	-6.3159	-3.4183

Table 5: Unstandardized canonical discriminant functions evaluated at group means

4.1.3 Canonical discriminant functions for multiple groups

Classification is performed by using feature variables and the 3 unstandardized canonical discriminant functions $\mathcal{UCDF}$. The coefficients of these functions can be seen in Table 4. The values of the $\mathcal{UCDF}$ evaluated at group means (cf. Table 5) give the following pattern of signs:

$$\begin{pmatrix} + & - & + \\ - & + & + \\ + & + & - \\ - & - & - \end{pmatrix} \rightsquigarrow \begin{pmatrix} \text{Scating} \\ \text{Fog} \\ \text{Holes} \\ \text{Scating + Fog} \end{pmatrix}$$

Of course, this is only a "rule of thumb" for classification, because the proper *cut–off* points obtained from the multivariate procedure must be used. To illustrate the possibility of real–time implementation of the algorithm, the compuation of the first discriminant function for the group mean of the first group – "Scating" – shall be demonstrated now:

$$\mathcal{UCDF}_{1,1} = \ell_{1,1,0} + \sum_{i=1}^{7} \ell_{1,1,i} \times x_i$$

where $\ell_{1,1,i}$ ($i = 0, 1, \dots 7$) are the coefficients of the $\mathcal{UCDF}_{1,1}$ and x_i ($i = 1, 2, \dots 7$) are the feature variables. Inserting the numerical values we obtain:

$$
\begin{aligned}
\mathcal{UCDF}_{1,1} = \\
-\ & 67.0162 \\
+\ & 2.2168 \times (-16.8968) \\
-\ & 2.4335 \times (-10.0484) + 45.6078 \times (0.8822) \\
+\ & 72.5525 \times (1.4222) - 32.0300 \times (2.0326) \\
-\ & 0.4794 \times (3.2028) + 0.2017 \times (21.9470)
\end{aligned}
$$

The result of these 7 multiplications and 8 additions is 1.186. Consequently, classification itself is very simple and can be performed in real–time even with moderate computational power.

Table 6 shows the results of prediction, i.e. classification of the observations 1, 7, 13 and 19. No error occured, 100% accuracy was achieved.

Obs.	Value of $\mathcal{UCDF}$ No.			Pred. Grp.
No.	1	2	3	
1	0.4690	-8.4919	15.2796	1$\surd$
7	-8.3681	4.2935	0.3135	2$\surd$
13	9.1518	6.3696	-0.4366	3$\surd$
19	-1.2544	-5.1226	-5.6608	4$\surd$

Table 6: Unstandardized canonical discriminant functions evaluated at the prediction set

4.1.4 Statistical inference

All results of statistical analyses have to be checked carefully for significance. Table 8 shows therefore a listing in extracts produced by the software package MDAD – *Multivariate Design, Analysis and Diagnosing* which was developed in the research group of the authors. This listing clearly indicates the overall significance of the discriminant analy-

sis, univariately as well as multivariately.

Table 7 gives an impression concerning the discriminatory power of the single feature variables for the overall performance. Table 9 shows the power of the single feature variables for the discrimination of the single groups.

A main point is, that *all* feature variables are needed for the discrimination of the 4 error pattern, even if some of them are not significant for the discrimination of certain groups.

Values of mean $\mathcal{SCDF}$–Coefficients for all feature variables						
Q	Ave	$ADev$	$SDev$	$SVar$	$Skew$	$Kurt$
1	2	3	4	5	6	7
1.1959	1.7329	2.6932	7.1225	8.1873	0.6406	1.8474

Table 7: Discrimination power of the feature variables referring to all $\mathcal{UCDF}$
High values denote a high discrimination power

4.2 Artificial Neural Networks

4.2.1 Introduction

Neural Networks[11] have arisen from analogies with models of the way that humans might solve the pattern recognition task. However, they have developed a long way from the biological roots. Great claims have been made for artificial neural networks, few of them have withstood careful checks. But, nevertheless neural networks have influenced pattern recognition practice very strongly. Although there is no full theoretical understanding of how they work at the moment, dozens of software packages and hardware accelerator cards have emerged on the market and disseminate the ideas of pattern recognition via artificial neural networks .

The most widely used and easily recognized neural network is the so-called *feedforward network*. By some estimates, it accounts for 90 percent of all neural network applications.
Typical applications outside the area of pattern recognition are for example: predictions of the stock market, recognition of handwriting, synthesis of speech from text, driving a truck.

4.2.2 Feedforward Network Operation

Figure 5 shows a common three–layer feedforward architecture. By "three–layer" is meant that the network has an *input layer*, one *hidden layer*, and an *output layer*. More hidden layers can be used. In general, three layers are sufficient to enable this type of network to model any *deterministic process* to a certain degree, therefore we will use it here for the purpose of pattern recognition.

Figure 5 actually represents a function acting on a vector, where the neural network function sends the vector $(x_1, ..., x_N)$ in $\mathbf{R}^N$ to the vector $(y_1, ..., y_M)$ in $\mathbf{R}^M$. Thus, the feedforward network can be represented as:

$$\mathbf{y} = F(\mathbf{x})$$

where $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_M)$. The action of this function is determined in a specific way. For a network with N input nodes, H hidden layer nodes, and M output nodes, the values y_k are given by:

$$y_k = g \left[\sum_{j=1}^{H} w_{jk}^O h_j \right]$$

with $k = 1 \ldots M$.
Variable w_{jk}^O is the output "weight" from hidden node j to output node k, and g is a function mapping $\mathbf{R}^1$ to $\mathbf{R}^1$. The values of the hidden layer nodes $h_j, j = 1, \ldots, H$ can be computed as follows:

$$h_j = \sigma \left[\sum_{i=1}^{N} w_{ij}^I x_i + w_j^T \right]$$

with $j = 1, \ldots, H$. Here, w_{ij}^I is the input "weight" from input node i to hidden node j, w_j^T is a threshold "weight" from an input node which has the constant value 1 to hidden node j , and x_i is the value at

```
          Multivariate Statistics - Discriminant Analysis
=======================================================================
     Number of variables: 7  Number of groups: 4  Number of cases:20
     Number of cases in groups: 5 5 5 5
     Group prior probabilities: 0.25  0.25  0.25  0.25
     Number of unknown observations to classify: 4
     Degrees of freedom due to hypothesis/error (nh)/(ne) : 3/16


     Univariate Wilks' Lambda and F-Ratio
     ------------------------------------
     gives the discrimination capability of the variables via
     variance shares.
               Alpha level:     0.01     (0.05)     [0.1]
               F-Table value:   5.2922   (3.2389)   [2.4618]
     The Chisquare statistics was transformed into a F-ratio.
     A variable is significant if the F-ratio > F-table value(3,16,alpha).
               Var.   Wilks     F-Ratio      Significance
               No.    Lambda
               ------------------------------------------------
                1     0.07634    64.53       yes   (yes)   [yes]
                2     0.05281    95.65       yes   (yes)   [yes]
                3     0.07685    64.07       yes   (yes)   [yes]
                4     0.11681    40.33       yes   (yes)   [yes]
                5     0.14571    31.27       yes   (yes)   [yes]
                6     0.15334    29.45       yes   (yes)   [yes]
                7     0.19623    21.85       yes   (yes)   [yes]


     Multivariate Test Criteria
     --------------------------
     If the Trace-criterion >= c * F-table value(3,16,0.01) it is
     significant-> Trace Criterion: 113.5858/Table value: 12.8234


     Canonical Discriminant Functions
     --------------------------------
     The unstandardized canonical discriminant functions (UCDF)
     are evaluated from the non-zero eigenvalues. An UCDF is
     significant if the Chisquare statistics > Chisquare-table
     value(DF,0.01).
     After  R.Wilks   Degrees of   Chisquare    Table      Signi-
     Fct    Lambda    Freedom      Value        Value      ficance
     ----------------------------------------------------------------
       1    0.00004      21        137.37589    38.93217     yes
       2    0.00267      12         80.01103    26.21696     yes
       3    0.09646       5         31.57141    15.08627     yes


     0 discriminant function(s) excluded from further analysis.
     3 discriminant function(s) remained after the test of
     significance.
```

Table 8: Listing in extracts

Group A	Group B	Var. No.	t-Statistics α–Level	1%	5%	10%
1	2	1	10.0669	yes	yes	yes
1	2	2	12.6259	yes	yes	yes
1	2	3	4.0255	yes	yes	yes
1	2	4	5.0988	yes	yes	yes
1	2	5	4.1253	yes	yes	yes
1	2	6	1.9012	**no**	**no**	yes
1	2	7	2.1328	**no**	yes	yes
1	3	1	11.3433	yes	yes	yes
1	3	2	11.9620	yes	yes	yes
1	3	3	6.6115	yes	yes	yes
1	3	4	3.2958	yes	yes	yes
1	3	5	3.3411	yes	yes	yes
1	3	6	2.6903	**no**	yes	yes
1	3	7	1.6284	**no**	**no**	**no**
1	4	1	6.5868	yes	yes	yes
1	4	2	3.9061	yes	yes	yes
1	4	3	4.1968	yes	yes	yes
1	4	4	4.7020	yes	yes	yes
1	4	5	3.8293	yes	yes	yes
1	4	6	5.4751	yes	yes	yes
1	4	7	5.2143	yes	yes	yes
2	3	1	1.2764	**no**	**no**	**no**
2	3	2	0.6639	**no**	**no**	**no**
2	3	3	10.6370	yes	yes	yes
2	3	4	8.3946	yes	yes	yes
2	3	5	7.4664	yes	yes	yes
2	3	6	4.5914	yes	yes	yes
2	3	7	3.7612	yes	yes	yes
2	4	1	3.4801	yes	yes	yes
2	4	2	8.7198	yes	yes	yes
2	4	3	0.1713	**no**	**no**	**no**
2	4	4	0.3968	**no**	**no**	**no**
2	4	5	0.2959	**no**	**no**	**no**
2	4	6	3.5739	yes	yes	yes
2	4	7	3.0815	yes	yes	yes
3	4	1	4.7565	yes	yes	yes
3	4	2	8.0559	yes	yes	yes
3	4	3	10.8083	yes	yes	yes
3	4	4	7.9978	yes	yes	yes
3	4	5	7.1705	yes	yes	yes
3	4	6	8.1653	yes	yes	yes
3	4	7	6.8428	yes	yes	yes

Table 9: Univariate t-Statistics estimates how two groups differ with respect to one variable. Alpha level/t-Table value: 0.01 0.05 0.10/2.92 2.12 1.75 (The local null hypothesis is rejected if the t-statistics > t-table value.)

input node i. Variable σ is the "sigmoid" function given by

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

and is called the activation function of the neural network.

Function g may be the same as the activation function or may be a different function. It is important that the activation function is nonlinear and has bounded output.

The action of the feedforward network is determined by its architecture (how many input, hidden, and output nodes it has), and the values of the weights. The numbers of input and output nodes are determined by the application, i.e. how many features, how many groups are within the pattern recognition tasks.

The number of hidden nodes is a variable that can be adjusted by the user. This adjustment depends upon experience of the user, although various methods for setting the number of hidden nodes, or *pruning* away unnecessary nodes, have been proposed.

With the architecture given, it is then the weight values which determine how the network performs. The process of adjusting these weight values in order to solve a certain pattern recognition task is called the *training* of the network. The network *learns* as the weight values are being modified to achieve the training goal.

However, the weights are nothing more than a set of parameters, determining the behavior of a particular function.

4.2.3 Backpropagation Training Algorithm

The starting point for any training procedure is data.

Training data consists of input–output pairs that have been generated by the process which the network is to emulate. For specific inputs, the feedforward network will produce its own set of outputs. The difference between the network outputs and the actual desired outputs is the "error" produced by the network. While training the network, we want to reduce this error to as small a value as

possible. Suppose, for simplicity, a single data point $(\mathbf{x}, \mathbf{d})$ consisting of an input vector $\mathbf{x} = (x_1, \ldots, x_N)$ and an desired output vector $\mathbf{d} = (d_1, \ldots, d_M)$. For a given set $\mathbf{W}$ of weight values, the feedforward network produces the output vector $y(\mathbf{W}) = (y_1(\mathbf{W}), \ldots, y_M(\mathbf{W}))$. One way to express the error $e(\mathbf{W})$ is :

$$e(\mathbf{W}) = \frac{1}{2} \sum_{k=1}^{M} (d_k - y_k(\mathbf{W}))^2$$

For a good performance of the feedforward network the weights $\mathbf{W}$ have to be chosen in such a manner that the following equation is valid:

$$e(\mathbf{W}) \to Min.$$

Notice that the actual value of $Min.$ depends upon the architecture of the network. Part of the art of training is determining the smallest number of hidden nodes that will do the job best, i.e. with smallest value of $Min.$

We have here a problem of convex optimization which can be solved numerically by the steepest descent algorithm. The iteration step can be written as:

$$\mathbf{W}_{new} = \mathbf{W}_{old} + \gamma \times (-grad(e(\mathbf{W})))$$

Variable $\mathbf{W}$ is a vector with $(N+1)H + HM$ components, $grad$ denotes the gradient with respect to $\mathbf{W}$, and γ is a the stepsize that controls the magnitude of the change in each iteration.

Backpropagation training is nothing more than the steepest descent algorithm applied to the vector of weight values and the error function.

The notion of "backward propagation" comes from the rather complicated expressions that arise when one actually computes the components of $grad(e(\mathbf{W}))$ by applying the chain rule for derivatives numerous times.

Unfortunately, steepest descent is a notoriously slow algorithm, requiring many iterations to reach an acceptable solution. Therefore, many improvements have been suggested, but they are beyond our scope here.
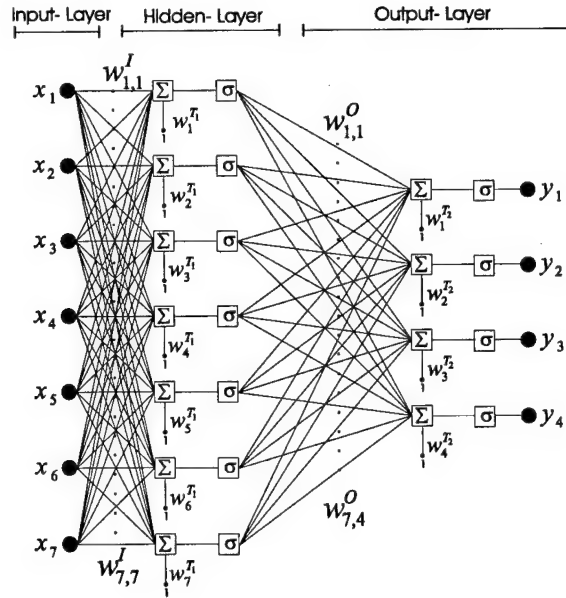
Figure 5: Structure of the three layer neural network

4.2.4 Example

Here we deal with the classification of the data set given in section 2.2 using a feedforward network. The neural network will be trained using the given data set with a simple rule:

> *if* **x** *is in the* i*th group* **then** *the* i*th component of* **y** *is one and the other components are zero.*

The output vector **y** is in $\mathbf{R}^4$ and the given input vector **x** is in $\mathbf{R}^7$. The neural network function will be choosen as

$$\mathbf{y}^T = \sigma(\mathbf{w}^O \cdot \sigma(\mathbf{w}^I \cdot \mathbf{x}^T + \mathbf{b}^{T_1}) + \mathbf{b}^{T_2}).$$

The σ function normalizes the range of output values to the range $\langle 0, 1\rangle$.
Consequently, the output activities **y** are easy to interpret. Using the backpropagation training algorithm with normalized random distributed initial conditions the weights can be found as:

$$w^I =$$
$$\begin{pmatrix} -1.3 & -0.8 & -0.1 & +1.2 & -3.0 & -0.3 & -0.6 \\ +1.9 & -5.9 & -1.3 & -4.3 & -3.7 & +4.4 & -1.0 \\ -3.1 & +5.7 & +4.3 & -3.6 & -2.1 & +3.5 & -0.5 \\ -4.4 & +2.8 & -7.6 & -4.9 & -5.9 & -4.1 & -0.3 \\ -1.5 & +3.0 & +2.7 & +5.7 & +2.9 & +1.6 & +0.9 \\ +2.2 & -1.8 & +6.5 & -2.0 & +3.0 & -2.6 & +0.3 \\ -1.5 & -1.7 & -2.2 & -2.7 & -1.1 & +0.8 & -0.7 \end{pmatrix}$$

$$w^O =$$
$$\begin{pmatrix} +1.0 & +1.9 & -19.8 & +12.9 & -3.0 & -3.0 & +6.9 \\ +0.0 & -5.7 & +5.5 & +7.6 & -0.3 & -10.4 & -8.6 \\ -0.9 & +5.2 & +17.6 & -3.1 & -16.4 & +12.3 & +5.0 \\ -0.4 & +4.4 & -3.3 & -13.4 & -0.3 & +12.2 & -9.0 \end{pmatrix}$$

$$b^{T_1} = \begin{pmatrix} -23.4 \\ +17.2 \\ +1.2 \\ +3.2 \\ -19.4 \\ +13.1 \\ -14.0 \end{pmatrix} \quad b^{T_2} = \begin{pmatrix} -6.5 \\ -0.3 \\ -9.7 \\ -1.6 \end{pmatrix}$$

Figure 6 shows the classified recognition set. The whole data set is correctly classified. The output activities for the prediction set are:

$$\mathbf{y_1} = \begin{pmatrix} 1.00 \\ 0.00 \\ 0.00 \\ 0.00 \end{pmatrix} \quad \mathbf{y_2} = \begin{pmatrix} 0.00 \\ 1.00 \\ 0.00 \\ 0.00 \end{pmatrix}$$

$$\mathbf{y_3} = \begin{pmatrix} 0.00 \\ 0.00 \\ 0.99 \\ 0.00 \end{pmatrix} \quad \mathbf{y_4} = \begin{pmatrix} 0.00 \\ 0.00 \\ 0.00 \\ 1.00 \end{pmatrix}$$

which is completely correct.
Therefore the neural network function found by the learning step represents a reliable pattern recognition model for this process.
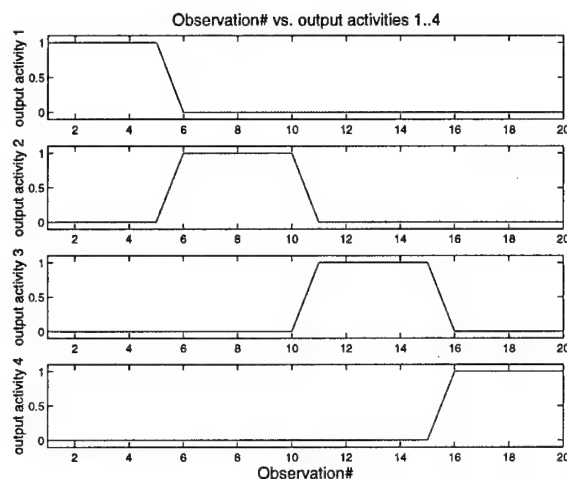
Figure 6: Observations vs. output activities y

4.3 Fuzzy Methods

4.3.1 Introduction

Fuzzy logic[11] is a means of dealing with information in the same way that humans or animals do, i.e. in imprecise terms.

Fuzzy logic is built around the concept of reasoning in degrees, rather than in boolean (yes/no, 0/1) expressions like computers do.

Variables are defined in terms of fuzzy sets. Rules are specified by logically combining fuzzy sets. The combination of fuzzy sets defined for input and output variables, together with a set of fuzzy rules that relate one or more input fuzzy sets to an output fuzzy set, comprise a fuzzy system.

Fuzzy systems represent well–defined static *deterministic* functions. Therefore reaction of a fuzzy system to inputs is anything but fuzzy. Inputs are presented to the system as specific values, and the fuzzy system produces a specific output value. The operation of a fuzzy system is thus analogous to that of conventional control systems.

Fuzzy systems are easy to design, typically require less computer power than alternative approaches, and provide robust operation.

However, since they can be regarded as an inductive approach, their use in pattern recognition applications can be a quite tricky one.

4.3.2 Fuzzy Sets

An ordinary set divides data into those items that are completely in the set and those items that are completely outside of the set.

We can describe this phenomenon by assigning the value 1 to all those data which are members of the set and the value 0 to all data which are not members of the set. For ordinary sets, only these two values are existent. The function which assigns these values is called the characteristic function of the set.

Fuzzy sets allow the possibility of degrees of membership. That is, any of the values between 0 and 1 (including 0 and 1) may be assigned. For example, given the fuzzy set "fast cars", we may speak of a particular car being a member of this set to degree 0.8. This would be a rather fast car, but not the fastest car imaginable.

The function which assigns this value is called the *membership function* associated with the *fuzzy set*. Fuzzy membership functions are the mechanism through which the fuzzy system interacts with the outside world. The range, or possible output values, of a membership function is the interval $[0, 1]$, the set of all real numbers between 0 and 1, inclusive. A typical choice for a fuzzy membership function is a piecewise linear trapezoidal function.

4.3.3 Fuzzy Rules

Fuzzy rules combine two or more input fuzzy sets, called the *antecedent* sets, and associate with them an output, or *consequent*, set. The antecedent sets are combined by means of operators that are analogous to the usual logical conjunctives "and," "or," etc.

One method of storing and representing fuzzy rules is through the use of a fuzzy associative memory (FAM) matrix. Figure 7 shows an example of a FAM matrix. In this example there are two inputs, X and Y. Each input variable has three fuzzy sets associated with it, which are labeled N, ZE, and P, for "Negative," "Zero," and "Positive" (these need not be the same for each input variable).

The output variable Z has five fuzzy sets associated with it: NL (Negative Large), NS (Negative Small), ZE (Zero), PS (Positive Small), and PL (Positive Large). Each FAM matrix entry is an output fuzzy set that is the consequent of a fuzzy rule. For example, the shaded entry in Figure 7 represents the rule:

> If X is Positive (P) and Y is Zero
> (ZE), then Z is Positive Small (PS).

("Zero" is a fuzzy set that would typically represent a range of values near 0.)

FAM matrices can have also dimensions higher than two. The number of inputs, or antecedents, to the fuzzy rules determines the dimension. Three inputs would result in a FAM matrix that looks like a three-dimensional cube. Higher numbers of inputs produce Hyper–FAM matrices.

In general it is not necessary to combine all system input variables into all of the fuzzy rules. However, this depends upon the structure of the system.

4.3.4 Fuzzy System Operation

Now there are fuzzy sets for the input and output variables, as well as fuzzy rules. How can a *defuzzified* output be obtained from a fuzzy system.? A simple method will be demonstrated at the system associated with the FAM matrix shown in Figure 7. In this case, there are two input variables, X and Y, with associated fuzzy sets N ("Negative"),ZE ("Zero"), and P ("Positive"). Figure 8 shows how the membership functions look for these sets.

The same membership functions will be used for X

and Y. However, this is in general not necessarily the case.

The membership functions for the fuzzy sets N, ZE, and P are denoted by F_N, F_{ZE}, and F_P, respectively. A particular value x of the input variable X then has membership degrees $F_N(x)$, $F_{ZE}(x)$, and $F_P(x)$.

For example, with the trapezoidal membership functions shown in Figure 8 and a value $x = 0.8$, we get:

$$F_N(0.8) = 0$$
$$F_{ZE}(0.8) = 0.7$$
$$F_P(0.8) = 0.2$$

The value $y = -0.7$ (cf. Figure 8) will give

$$F_N(-0.7) = 0.13$$
$$F_{ZE}(-0.7) = 0.8$$
$$F_P(-0.7) = 0.0$$

Now we present $x = 0.8$ and $y = -0.7$ to the system as values of the input variables X and Y. The idea is to assign a weight value to each entry in the FAM matrix by taking the minimum of the membership function values associated with that entry. Consider the FAM matrix entry corresponding to X a member of the fuzzy set ZE, and Y a member of the fuzzy set N (cf. Figure 9). The weight w_1 associated with this entry is:

$$\begin{aligned} w_1 &= min\{F_{ZE}(0.8), F_N(-0.7)\} \\ &= min\{0.7, 0.13\} \\ &= 0.13 \end{aligned}$$

Clearly, only those FAM matrix entries which have nonzero membership-function values for both X and Y will have nonzero weights associated with them.

In this sense it is said that the rules corresponding to these entries are *activated*.

The marked squares in Figure 9 show the four activated rules for the values in this example. In addi-

tion to w_1, there exist three more nonzero weights:

$$w_2 = min\{F_{ZE}(0.8), F_{ZE}(-0.7)\}$$
$$= min\{0.7, 0.8\}$$
$$= 0.7$$

$$w_3 = min\{F_P(0.8), F_N(-0.7)\}$$
$$= min\{0.2, 0.13\}$$
$$= 0.13$$

$$w_4 = min\{F_P(0.8), F_{ZE}(-0.7)\}$$
$$= min\{0.2, 0.8\}$$
$$= 0.2$$

The output variable Z consists of five fuzzy sets: NL, NS, ZE, PS, and PL.
Rather than treat these as fuzzy sets, we're going to assign specific values to them.
This does not imply that the fuzzy system output is limited to a finite number of values. This will result in the fact that NL, NS, ZE, PS, and PL now represent specific numerical values.
These values can be computed as follows:

$$Out = \frac{(w_1 NS + w_2 ZE + w_3 NS + w_4 PS)}{\sum_{i=1}^{4} w_i}$$

With more general output fuzzy sets, determination of the defuzzified output involves computation of centroid values of – in general – overlapping membership functions.

X

Figure 7: An example of a two-dimensional FAM matrix.
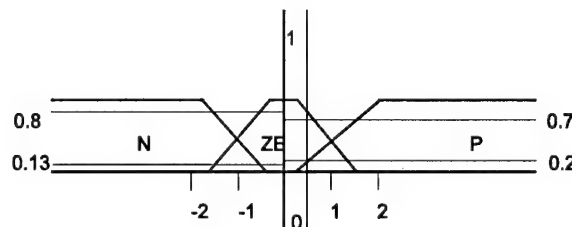
Figure 8: Fuzzy membership functions. The point 0.8 is a member of the fuzzy set P to degree 0.2, and a member of the fuzzy set ZE to degree 0.7. The point -0.7 is a member of the fuzzy set N to degree 0.13 and a member of the fuzzy set ZE to degree 0.8.

4.3.5 Example

Here we use a *Sugeno*–type fuzzy inference system.

The fuzzy system function will be adjusted in the same way as shown in the prcededing section:

X

		N	ZE	P
	N	NL	NS	NS
Y	ZE	NS	ZE	PS
	P	PS	PS	PL

Figure 9: FAM matrix. The shaded entries are the activated rules

if the input vector **x** *is in the* i*th group* *then the* i*th component of output vector* **y** *is one and the other components are zero.*

Concerning the given data set the output vector **y** is in $\mathbf{R}^4$ and the given input vector **x** is in $\mathbf{R}^7$. The first-order *Sugeno* fuzzy model has rules in the following form:

if x_1 *is* A_1 *and .. and* x_7 *is* B_7 *then* $z = p_1 \cdot x_1 + .. + p_7 \cdot x_7 + r,$

where x_i are the components of an input vector **x**, A_i are the fuzzy sets, p_i and r are constants. Five fuzzy rules will be used to model the fuzzy input. As membership functions the *Gaussian* curve $f(x) = e^{(-(x-x_0)^2/(2\sigma^2))}$ will be employed. The AND linkage will be defined as logical product. The consequent term of the used rule includes four output functions. Each output function is a weighted linear combination of the fuzzy input. The defuzzification is simply a weighted and normalizied average of the output functions using the activation functions of each rule. All constants will be calculated with the least squares method.

In result, each input vector (recognition set) is correctly classified. Therefore, learning of the fuzzy system worked quite well.

However, with the prediction set as input for the

classification system we obtain

$$y_1 = \begin{pmatrix} -0.98 \\ -0.01 \\ 0.00 \\ -0.02 \end{pmatrix}, \quad y_2 = \begin{pmatrix} -0.98 \\ 1.00 \\ 0.00 \\ -0.00 \end{pmatrix}$$

$$y_3 = \begin{pmatrix} 0.03 \\ 0.00 \\ 1.00 \\ 0.00 \end{pmatrix}, \quad y_4 = \begin{pmatrix} 3.17 \\ 0.00 \\ -0.00 \\ 1.02 \end{pmatrix}$$

which is only correct for pattern 3.

This shows that the fuzzy system function does not represent a reliable pattern recognition model for this process.

The fuzzy inference system studied here can therefore only be a starting step to build a robust fuzzy classifier. The logical linkages of each rule, the number of rules used and the type of membership functions must be changed to achieve correct pattern recognition. The optimization process will need certainly many iterations because the number of degrees of freedom is high.

This simple example highlights the difficulties and limitations in using and optimizing fuzzy inference systems.

4.4 Neuro–Fuzzy Methods[12]

4.4.1 How it works ...

The main advantage of fuzzy systems is that a desired system behaviour can be implemented by simple reasoning operations. This will often provide a solution with a small effort in terms of cost, time, and manpower. Additionally, all engineering

knowledge can be used to improve system performance.

However, this advantage is also a limitation, because, like in pattern recognition, knowledge has to be extracted *manually* from data. This is a problem especially with large and noisy data sets.

On contrary, because a neural net can be trained from data, it provides a – more or less – good solution for a given task without any help from humans. Despite this, only a few applications exist for neural networks in practice. This is mainly due to the large amount of computing power needed to solve even moderate sized classification tasks. Furthermore, a neural net remains a black box, since it can not be interpreted what certain weights may mean.

This is the reason, why the idea arose to establish a hybrid system, namely a combination of a neural net and a fuzzy logic system. So neuro–fuzzy systems were developed which combine the explicit knowledge representation of fuzzy logic with the learning power of neural nets.

The training of neuro–fuzzy systems is rather complicated since the backpropagation algorithm can not be used, because it needs explicit differentiation of the activation functions of the neurons. This is, however, not possible with fuzzy logic rules, because certain mathematical functions like minimum and maximum are used.

The most common solution to this problem is the use of FAM (fuzzy associative memory) matrices. Algorithms were developed which map FAM's to neurons.

Therefore error backpropagation methods with neuro fuzzy systems can be used.

4.4.2 Example

Last, but not least we consider the Adaptive Network based Fuzzy Inference System (ANFIS).
ANFIS is used as learning algorithm for the *Takagi–Sugeno* system which was introduced already in section 4.3.5.
The *Takagi– Sugeno* fuzzy system can be rewritten as a feedforward neural network.
Figure 10 shows it in layer structure form. The fuzzy system function is given as:

if x_1 *in* A_1 *AND* x_2 *in* B_1 *then*
$$y_1 = c_{1,0} + c_{1,1}x_1 + c_{1,2}x_2$$

if x_1 *in* A_2 *AND* x_2 *in* B_2 *then*
$$y_2 = c_{2,0} + c_{2,1}x_1 + c_{2,2}x_2$$

with the defuzzification

$$y = \frac{\beta_1 y_1 + \beta_2 y_2}{\beta_1 \beta_2}$$

when the activation functions of each rule will be denoted as β_1 and β_2.
The layer L_5 can be defined as a single layer *Adaline* network with variable weights.
The activation functions of each rule are the variable weights. The parameters can be adjusted using a backpropagation algorithm alone, or in combination with a least squares method.
The learning rule will be in our case:

if the input vector $\mathbf{x}$ *is in the* i*th group*
then the $\mathbf{y}$ *presents the value* i.

The group numbers are here $0 \ldots 3$.
Figure 11 shows the used structure. This structure consists of 18 rules with gausssian membership functions and linear output functions.
Thie structure was trained with an hybrid algorithm of backpropagation and the least squares method. Figure 12 shows the – correct – classification results of the recognition set.
Also all members of the prediction set were classified correctly.
Therefore the neuro fuzzy system function is robust and gives a perfect pattern recognition system with our small example.

$$y_1 = 0.00, \quad y_2 = 1.01$$
$$y_3 = 1.97, \quad y_4 = 2.92$$

5   Conclusions

A short introduction to pattern recognition has been presented. It has been shown that powerful methods exist, however, care has to be taken to build robust and consistent classifiers. The best approach for the unexperienced user seems to be the use of classical statistical tools, since *plug and play* works in this case. Otherwise a profound background knowledge on the behaviour of the methods used is needed, since mis–classification or overfitting may occur without notice.

6 References

1. *C. M. Bishop*, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.

2. *G. X. Ritter and J. N. Wilson*, Computer Vision Algorithms in Image Algebra, CRC Press, Boca Raton, New York, Tokyo, London, 1996.

3. *J. C. Stover*, Optical Scattering: Measurement and Analysis, Optical and electrooptical engineering series, Mc. Graw–Hill, 1990.

4. *P. Roche and E. Pelletier*, "Characterization of optical surfaces by measurement of scattering distribution", Applied Optics Vol. 23 (1984), pp. 3561–3566

5. *H. Rothe and H. Truckenbrodt*, "Discrimination of Surface Properties Using BRDF–Variance Estimators as Feature Variables", Proc. SPIE No. 1781 (1992), pp. 152–162.

6. *J. E. Jackson*, A Users Guide to Principal Components, John Wiley & Sons, New York, 1991.

7. *H. Rothe, A. Duparré*, P. Riedel and M. Timm, "Real–time defect detection using multi–aperture fiber optic sensors and machine learning", Proc. SPIE No. 1989 (1993).

8. *H. Rothe, O. Ginter and C. Woldenga*, "Assessment and robust reconstruction of laser radar signals", Optics and Laser Technology, 25(5), 289–297 (1993).

9. *P. A. Lachenbruch*, Discriminant Analysis, Hafner Press, New York, 1974.

10. *W. R. Dillon and M. Goldstein*, Multivariate Analysis, John Wiley & Sons, New York, 1984.

11. *S. T. Welstead*, Neural Network and Fuzzy Logic Applications in C/C++, John Wiley & Sons, New York, 1994.

12. *C. von Altrock*, Fuzzy Logic & NeuroFuzzy Applications explained, Prentice Hall PTR, Upper Saddle River, New Jersey 07458, 1995.
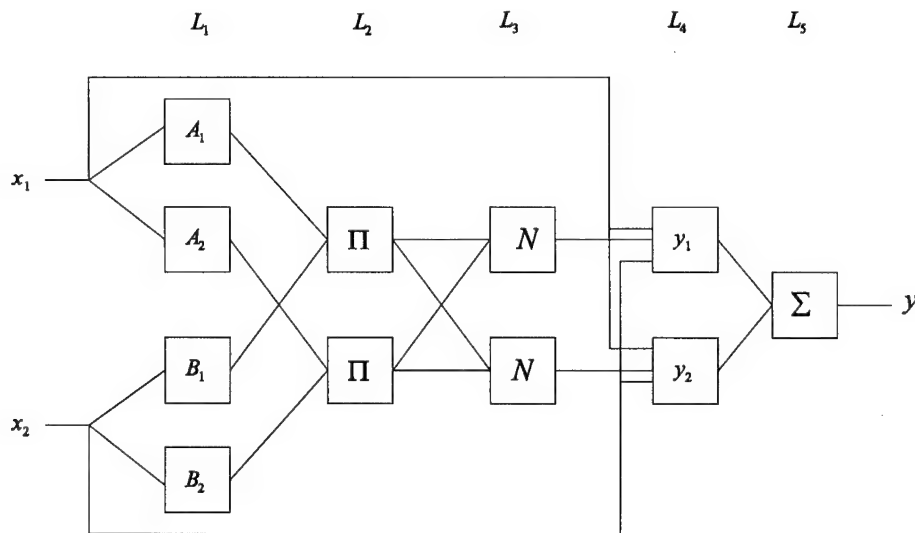
Figure 10: ANFIS - simple structure
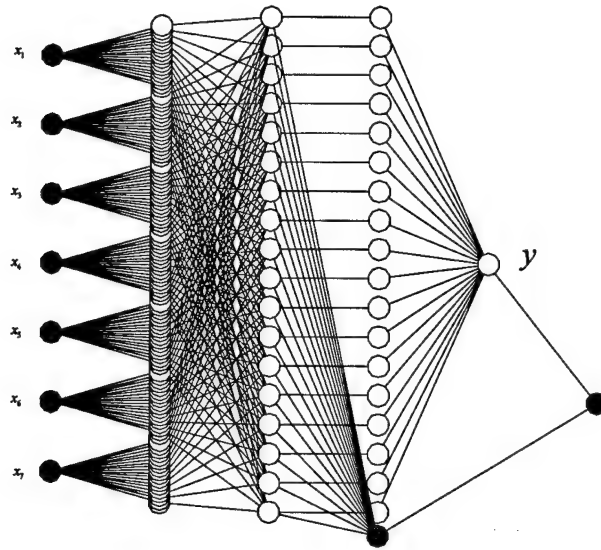
Figure 11: ANFIS - example structure
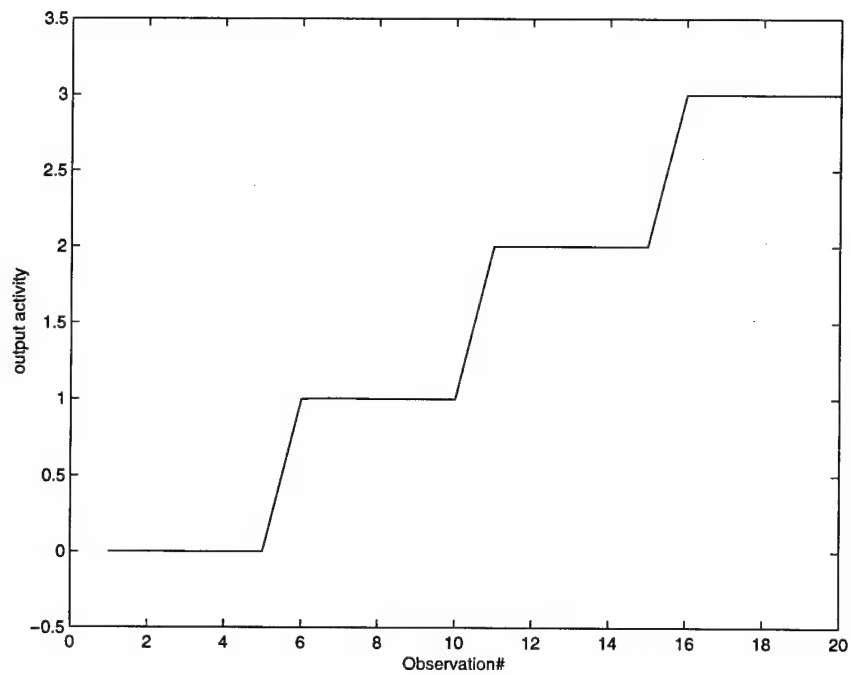
Figure 12: Observations vs. output activity

Signature Based Target Recognition

Carl E. Baum
Air Force Research Laboratory
AFRL/DEHP, Building 909
3550 Aberdeen Avenue S. E.
Kirtland AFB, NM 87117-5776
U. S. A.

1. SUMMARY

In identifying a target as a member of a target class (e.g., a particular model of aircraft) one can use the scattering signatures from a transient type of radar (or a radar with a large number of frequencies). There are various types of signatures that one can use and a number of the most common ones are discussed.

2. INTRODUCTION

One way to identify a target is to "look" at it in the sense of a photograph so that one observes shape, color, etc., as the relevant items for identification. This is often referred to as imaging and implies wavelengths short compared to dimensions of interest on the target, and quite small angular resolution if the target is far away. This is an important approach to target recognition, but not the approach considered here. Another approach is called inverse scattering, but this is quite complicated and is also not considered here.

3. SIGNATURES

In this paper we consider identification by means of signatures. For this we consider model-based parameters. By a scattering model we mean some mathematical expression with a not-too-large set of parameters which represent the scattering (exact or approximate) over some region of time, frequency, etc. The parameters may be aspect dependent or aspect independent (more desirable). So we can define

signature type: a set of parameters associated with a scattering model

signature: a set of specific parameter values (including aspect dependence) associated with a signature type and related scattering model.

A target may have more than one signature due to the applicability of more than one scattering model.

4. SCATTERING MODELS

There are various scattering models of interest including

1. Singularity expansion method (SEM)
    – poles (natural frequencies) and residues

2. Generalized cone (dilation symmetry)

3. High-frequency method (HFM)
    – asymptotic

4. Linear array of scatterers

5. Scattering centers

6. Low-frequency method (LFM)
    – electric and magnetic polarizabilities

A unifying concept for these models is symmetry, either of the whole target or of some of its substructures (combined with temporal isolation (causality)).

5. CONCLUDING REMARKS

There is already a large literature on this subject. The accompanying bibliography emphasizes the larger review papers and book chapters where an enormous number of references (not repeated here) can be found. One may think of this as a metabibliography.

BIBLIOGRAPHY

General

1. D. G. Dudey and D. M. Boodman, Transient Identification and Object Classification, ch. 13, pp. 456–497, in E. K. Miller (ed.), *Time-Domain Measurements in Electromagnetics*, Van Nostrand Reinhold, 1986.

2. C. E. Baum, Signature-Based Target Identification and Pattern Recognition, IEEE Antennas and Propagation Magazine, Vol. 36, No. 3, June 1994, pp. 44–51.

3. C. E. Baum, Concepts in Transient/ Broadband Electromagnetic Target Identification, pp. 515–525, in L. Carin and L. B. Felsen, (eds.) *Ultra-Wideband, Short-Pulse Electromagnetics 2*, Plenum Press, 1995.

4. C. E. Baum, Symmetry and Transforms of Waveforms and Waveform Spectra in Target Identification, ch. 7, pp. 309–343, in C. E. Baum and H. N. Kritikos (eds.), *Electromagnetic Symmetry*, Taylor & Francis, 1995.

5. C. E. Baum, Decomposition of the Backscattering Dyadic, pp. 155–173, in H. Mott and W.-M. Boerner (eds.), Wideband Interferometric Sensing and Imaging Polarimetry, Proc. SPIE, Vol. 3120, 1997.

6. C. E. Baum, Target Symmetry and the Scattering Dyadic, in D. H. Werner and R. Mittra (eds.), *Frontiers of Mathematical Methods in Electromagnetics*, IEEE Press (in publication).

General references on singularity expansion method (SEM)

7. C. E. Baum, Emerging Technology for Transient and Broad-Band Analysis and Synthesis of Antennas and Scatterers, Proc. IEEE, pp. 1598–1616, 1976.

8. C. E. Baum, The Singularity Expansion Method, ch. 3, pp. 129–179, in L. B. Felsen (ed.), *Transient Electromagnetic Fields*, Springer-Verlag, 1976.

9. C. L. Dolph and R. A. Scott, Recent Developments in the Use of Complex Singularities in Electromagnetic Theory and Elastic Wave Scattering, ch. 14, pp. 503–570, in P. L. E. Uslenghi (ed.), *Electromagnetic Scattering*, Academic Press, 1978.

10. C. E. Baum, Toward an Engineering Theory of Electromagnetic Scattering: The Singularity and Eigenmode Expansion Methods, ch. 15, pp. 571–651, in P. L. E. Uslenghi (ed.) *Electromagnetic Scattering*, Academic Press, 1978.

11. L. W. Pearson and L. Marin (eds.), Special Issue on the Singularity Expansion Method, Electromagnetics, 1981, pp. 349–511.

12. K. J. Langenberg (ed.), Special Issue on Transient Fields, Wave Motion, Vol. 5, 1983, pp. 297–411.

13. C. E. Baum, The Singularity Expansion Method: Background and Developments, IEEE Antennas and Propagation Newsletter/Magazine, Vol. 28, No. 4, August 1986, pp. 15–23.

14. C. E. Baum, E. J. Rothwell, K.-M. Chen, and D. P. Nyquist, The Singularity Expansion Method and Its Application to Target Identification, Proc. IEEE, 1991, pp. 1481–1492.

15. E. K. Miller, Model-Based Parameter-Estimation Applications in Electromagnetics, ch. 12, pp. 205–256, in B. deNeumann (ed.), *Electromagnetic Modelling and Measurements for Analysis and Synthesis Problems*, Kluwer Academic Publishers, 1991.

16. C. E. Baum, SEM and EEM Scattering Matrices and Time-Domain Scatterer Polarization in the Scattering Residue Matrix, ch. I-9, pp. 427–486, in W.-M. Boerner et al (eds.), *Direct and Inverse Methods in Radar Polarimetry*, Kluwer Academic Publishers, 1992.

17. H. Überall (ed.), *Acoustic Resonance Scattering*, Gordon and Breach Science Publishers, 1992.

18. H. Überall, Fine Resolution of Radar Targets, ch. 3, pp. 47–112, in W.-M. Boerner and H. Überall (eds.), *Radar Target Imaging*, Springer-Verlag, 1994.

19. T. A. Sarkar and O. Pereira, Using the Matrix Pencil Method to Estimate the Parameters of a Sum of Complex Exponentials, IEEE Antennas and Propagation Magazine, Vol. 37, No. 1, February 1995, pp. 48–55.

20. C. E. Baum, Direct Construction of a ξ-Pulse from Natural Frequencies and Evaluation of the Late-Time Residuals, Interaction Note 519, May 1996.

21. C. E. Baum, Properties of Eigenterms of the Impedance Integral Equation, ch. 2, pp. 39–91, in A. Guran, R. Mittra, and P. J. Moser (eds.), *Electromagnetic Wave Interactions*, World Scientific, 1996.

22. C. E. Baum, Representation of Surface Current Density and Far Scattering in EEM and SEM with Entire Functions, in P. P. Delsanto and A. W. Saenz (eds.), *New Perspectives on Problems in Classical and Quantum Physics*, Gordon and Breach Science Publishers (in publication).

Early SEM papers

23. C. E. Baum, On the Singularity Expansion Method for the Solution of Electromagnetic Interaction Problems, Interaction Note 88, December 1971.

24. L. Marin and R. Latham, Analytical Properties of the Field Scattered by a Perfectly Conducting, Finite Body, Interaction Note 92, January 1972; also L. Marin, Natural-Mode Representation of Transient Scattered Fields, IEEE Trans. Antennas and Propagation, 1973, pp. 809–818.

25. F. M. Tesche, On the Singularity Expansion Method as Applied to Electro-magnetic Scattering from Thin-Wires, Sensor and Simulation Note 102, April 1972; also F. M. Tesche, On the Analysis of Scattering and Antenna Problems Using the Singularity Expansion Technique, IEEE Trans. Antennas and Propagation, 1973, pp. 53–62.

Scattering centers

26. C. Ray Smith and P. M. Goggans, Radar Target Identification, IEEE Antennas and Propagation Magazine, Vol. 35, No. 2, April 1993, pp. 27–38.

Generalized cone signature

27. C. E. Baum, Continuous Dilation Symmetry in Electromagnetic Scattering, ch. 3, pp. 143–183, in C. E. Baum and H. N. Kritikos, (eds.) *Electromagnetic Symmetry*, Taylor & Francis, 1995.

WAVELET TECHNIQUES AND OTHER MULTIRESOLUTION TECHNIQUES FOR TARGET PHENOMENOLOGY STUDIES

Eric K. Walton
The Ohio State University
Electrical Engineering Department, ElectroScience Laboratory
1320 Kinnear Road, Columbus, Ohio 43212-1191
PH   614 292 7981, FAX 614 292 7297
walton.1@osu.edu

1. SUMMARY

Many types of radar systems are available that can obtain the band limited calibrated impulse response of the radar target. This can be done by using a broad band coherent radar and then transforming from the frequency domain to the time domain, or it can be done by actually transmitting a very short time impulse (or other waveform) and receiving the voltage versus time response of the signal scattered from a target (the technique shown as an example here).

Considerable information about the radar target electromagnetic phenomenology can be obtained by time-frequency analysis of the resulting band limited impulse response. Such time-frequency analysis of the impulse response is usually at the upper limit of spectral resolution in cases where the target has resonant effects over the radar frequency band. In this paper, we will discuss these time frequency technniques especially including wavelet transformations and show examples for a specific target measured using an impulse radar.

2. INTRODUCTION

The ultra-wideband radar scattering from an object contains considerable information about the phenomenology of the target/electromagnetic interactions. Often, the data are measured as scattering amplitude and phase over a very wide band of frequency increments (the frequency domain). Recently, there has been great interest in applications involving transmission and reception of very short impulses (0.1 to 2 ns) (or other short-time/wide bandwidth waveform) to obtain a (band limited) impulse response of the target. The author has done conciderable recent research, for example, on a class of radar systems that transmit random electromagnetic noise and obtain the target (band limited) impulse as a cross correlation of the transmitted signal and the received signal.

It is possible to transform from one domain to another (FFT and IFFT), so these different measurement techniques are equivalent as long as the frequency band is equivalent. Extraction of target phenomenology or radar target identification is possible by processing the data in either the time or the frequency domain. On the other hand, it can be shown that most scattering objects will have different spectral characteristics at different propagations times. The early time scattering may be dominated by specular scattering (events narrow in time extent) and the late time may be dominated by resonant "ringing" (events narrow in frequency). In this sense, it is necessary to consider such radar targets as "non-stationary" or "dispersive" because the spectral response is a function of propagation time. One can obtain images (signatures or mappings) of the scattering behavior of the target by computing time-frequency representations of the scattering. We will discuss several algorithms for doing this here and show examples of the resulting response images.

For simplicity, a single example data set will be used for the discussions here. We will use a generic missile shape and a set of data files containing data for the scattering of a short pulse from a mock missile shape (a cylinder with a conic frustrum shape for a nose) provided by Mr. Jamie Henderson and Ms. J.Arthur [17]. In the measurement range used here [17], and shown in Figure 1, a short EM impulse is propagated over a ground plane to the missile shape. Scattering points on the mock-missile are labeled in the figure. As the pulse propagates, it is detected by a D-dot probe (a voltage probe that responds to the derivative of the actual E-field) before it illuminates the missile and also after it scatters from the missile (near-field bistatic). The scattered waveforms in this case were sampled at a rate of 0.02 ns per sample. The digitizer is able to sample the wave form as fast as 0.002 ns per sample. This provides information on the waveforms with a spectral span of 25 GHz. Spectral analysis of the data shows that the energy in the illumination impulse and the scattered signal is limited to less than 8 GHz, well within the capability of the measurement system.

The missile was constructed with a removable probe (a 5 cm rod) on the nose attached to a bulkhead coaxial connector. Depending on the test, the bulkhead connector was terminated inside the missile with an open circuit or a short circuit.

The missile orientation is shown in the figure. It is tilted at a 45 degree angle toward the radar. Note that there is a Styrofoam support block not shown here (which may

cause some spurious scattering). The data files discussed here and the configurations are summarized below.

Table 1. Missile configurations

FILE NAME	PROBE	INTERNAL
7USA3	NO	SHORT
6USA8	NO	OPEN
7USA2	YES	SHORT
7USA9	YES	OPEN

The missile is a cylinder/frustrum with a removable conductive probe on the nose. The cylinder is 118.70 cm long, the frustrum 14.6 cm long. The diameter of the cylinder and base of frustrum is 15.75 cm tapering to 8.5 cm. The probe was 5.35 cm long. The bulkhead coaxial connector at the tip may have a short circuit, open circuit or matched load attached. There was also a data file called IMPULSE2.DAT which is the signal radiated from the TEM antenna as sensed by the D-dot probe. It is approximately the time derivative of the field incident upon the scatterer and can be used for calibration.

The goal is to use time and frequency based signal processing techniques to extract signatures from the data which reveal the behavior of the time and frequency dependent response of the missile. A number of algorithms were implemented and used to analyze the data and are described in the next several sections of this paper.

3. EVALUATION OF POTENTIAL SCATTERING MECHANISMS

If we consider the possible scattering mechanisms listed below, we can estimate the propagation times and evaluate those regions on the time plots and in the time frequency studies that follow.

1. Source → top of frustrum → Ddot probe:
   Time = 0.0 ns (reference).
2. Source → top of frustrum → Ddot probe:
   Time = –2.0ns
3. Source → top of frustrum → base → Ddot probe:
   Time = 4.0 ns.
4. Source → top of frustrum → base → frustrum → Ddot probe: Time = 8.3 to 9.05 ns.

Also, if we consider the fully illuminated top of the frustrum, we see that it may scatter an attached wave traveling down the body of the missile. It will be a fairly strong term. Thus items 3 and 4 above are likely candidates. If we look at the actual measured data, we clearly see 1, 2 and 4. Thus in the actual data, the "base to D-dot probe" term is not seen. Finally, the ring-type discontinuity that the top of the frustrum presents will scatter energy toward the D-dot probe at a set of frequencies where the multiple paths (near side/far side

etc.) add in phase in the direction of the D-dot probe. An estimate, based only on path distances, gives enhanced scattering toward the D-dot probe at approximately 2, 4, 6 and 8 GHz. When we look at the frequency behavior of this mechanism using the time frequency distribution (TFD) analysis, we will see such spectral behavior for this term.

If there are effects due to the probe attached to the top of the frustrum, we expect to see probe induced changes in the frustrum top term, or resonances. The resonances may be due to the probe itself, or due to the cavity inside the missile body (if the bulkhead connector is open circuited). We can look for spectral effects near the probe scattering term. The inside of the missile body may also behave like a shorted waveguide. We may see dispersive scattering due the shorted end of this circular waveguide for the case of the probe with the open circuit. The time delay of this term from the direct probe scattering term will be dispersive but will be approximately the same as the external delay term discussed above (Number 4 at 8 to 9 ns from the direct top of frustrum term).

4. CALIBRATION

The scattering data measured by the D-dot probe is the derivative of the actual E-field. It is necessary to integrate the data in order to obtain a voltage waveform which is directly proportional to the E-field detected voltage. Even if this is done, the resulting E-field waveform contains time and frequency dispersive effects due to the signal source, antenna, detector (D-dot probe), coaxial transmission lines and sensor. We may wish to also remove these effects.

It would be possible to remove these dispersive effects if we had scattering data on a radar scatterer with a theoretically known RCS (and thus radar impulse response). To do that, we would take the raw data scattered from the target under test (TUT), $V_{tut}(t)$ and the raw data scattered from the calibration target $V_c(t)$ and integrate. This would give $h_{tut}(t)$ and $h_c(t)$, the voltage waveforms proportional to the scattered E-fields for the target under test and the calibration target. At this point, we would take the Fourier transform of $h_{tut}(t)$ and $h_c(t)$ to obtain $S_{tut}(w)$ and $S_c(w)$, the frequency spectra proportional to the scattered E-fields. Finally, we would compute $S_{cal}(w)$, the calibrated scattered field in units of $\sqrt{(area)}$ or $(\sqrt{RCS})$.

$$S_{cal}(\omega) = S_{exa}(\omega) \cdot \frac{S_{tut}(\omega)}{S_{cal}(\omega)}$$

We would have the square root of RCS (at this point a complex number) of the target as a function of frequency.

Now at this point, a large part of the calibrated spectrum would still be incorrect. This is because over a large part

of the spectrum, the spectral energy will be very low because the system can not produce or radiate energy in these frequency bands. Typically, only the band where the system efficiently produces and radiates energy will be valid. This region can be determined by integrating the D-dot data and transforming it to the frequency domain. Only the non-zero (approximately zero) region will be valid. Outside of this band, the denominator in the calibration equation is very small and the result is dominated by measurement noise. We set the unreliable low frequency data points to zero and throw away the unreliable high frequency data points. (Danger - Do not throw away the low frequency part of the array, we will need this string of zeros extending from the first useful frequency point to the DC value later). Finally, if we wish, we may take the Fourier transform of this final set of data to yield the calibrated impulse response of the target under test.

Unfortunately, in the available data set in this example, we do not have scattered data from a calibration reference. We do have data on the incident E-field as measured by the D-dot probe. If we assume that the data from the D-dot probe is flat, we can perform a transformation of the data which will permit the relative behavior of the scattered field to be determined. In this case, we integrate both the incident voltage waveform and the scattered waveform. We have done this, and some of the results will be shown in the next section.

Having said all this, it is important to make an observation about the Fourier transform. We know that the Fourier transform of the integral of any time series is simply $1/j\omega$ times the Fourier transform of that time series. Thus the process of taking the integral simply introduces a spectral roll-off in the result as a function of frequency. If we do this, we note that at 8 GHz the raw data PSD is 17 dB stronger than the integrated data PSD. Also, given the type of data measured here, I believe that we can trust the dynamic range to extend to approximately 35 dB. So, if -35 dB is used as a criteria, the raw data has an upper frequency limit of approximately 9.5 GHz, while the integrated waveform has an upper frequency limit of 8 GHz.

For the studies to be done here, rapid frequency rolloff is a disadvantage. We want to observe the spectral response of resonant structures on a test target (the missile shape). This integration process will actually suppress these effects. We have done studies both with and without integration and have shown that the resonant effects that we wish to see are more clearly revealed if the integration process is not done. Note also that the PSD for the raw data is more flat (has less rapid rolloff) than when the integration is done. For these reasons, we will show most of the spectral responses in the following section by processing only the raw data. The only penalty will be that the impulse response that we get will

not be calibrated. The higher frequencies will be enhanced in our work.

5. SHORT TIME FOURIER TRANSFORM ANALYSIS

5.1 The Short Time Fourier Transform Algorithm

Much of this work is concerned with the behavior of the spectrum of the mock missile as a function of time delay. This analysis will attempt to reveal the specific onset of resonance effects as is expected from a resonant scatterer. The analysis will consist of forms of time frequency representations (TFR's) or time frequency distributions (TFD's) [12, 14 to 20, 22 to 24]. These are representations or mappings of the spectral behavior of the target scattering as a function of time. The most intuitive of these algorithms is the short time Fourier transform (STFT). In this algorithm (in our case), the time domain data are broken into overlapping segments (a rectangular window) and each segment is then Fourier transformed into the frequency domain and the amplitude squared taken to form the power density versus frequency as a function of time. We then plot the data as a two dimensional color mapping with frequency vertical and time horizontal. The color scale for the image is scaled to the radar scattering power at each particular frequency and time. The algorithm is sketched in Figure 2.

We have applied this algorithm, and a data subtraction version (target with probe minus target without probe) to each of the data files. The results are given in the next section.

5.2 STFT Applied to the Tilted Missile

If we apply the STFT to the 45 degree tilt case, we can look for interaction terms and resonance terms at different times. We have data with and without the 5 cm. probe and for the cases of a short circuit on the connector (on the inside) and an open circuit.

Consider first the 45 deg. tilt missile with no probe and with a short circuit on the bulkhead connector inside the missile (7USA3) as shown in Figure 3. Note that a distinct response is found at approximately 2.3 ns, 7.3 ns and 11.0 ns. These responses correspond to STFT responses (shifted 2 ns earlier as mentioned above) as listed below

TIME	SPECTRAL RESPONSE	MECHANISM TIMING
0.3 ns	1.6 GHz response	top and side of frustrum
5.2 ns	2.4 to 0.4 GHz response	top to base to D-dot
9.0 ns	6.4 GHz to DC resp. (dispersive)	top to base to top to D-dot

The mechanisms were previously listed in section 3. Also note the small response at approximately 5.6 GHz at 0.3 ns (STFT). Perhaps this is a higher order term (double bounce) related to the strong 1.6 GHz effect at this time. The triple bounce term at 9.0 ns. in the STFT display is particularly interesting because of its dispersive nature (frequency response is function of delay time). Since there is no probe on the missile, none of this behavior is due to the probe.

Next, we applied the same algorithm to 7USA2, as shown in Figure 4. In this case, there is a probe on the tip of the missile, but the inside is terminated in a short circuit, so there can be no internal effects. We also applied the STFT algorithm to 6USA9 in Figure 5, where the inside of the bulkhead connector is open circuited. In this case, there is an opportunity for some signal propagation and resonances due to the inside of the missile case. Unfortunately, there are no obvious differences in these three figures. The addition of the probe causes no strong effect in the data.

We also applied the STFT algorithm to the response data after integration. The result should be more truthful in the sense that the roll-off in frequency should be included. The result, shown in Figure 6 for 6USA9 (with probe and open circuit), shows nearly equal spectral energy levels at the three times identified here. The frequency coverage is also nearly the same (approx. 5 GHz to DC as seen in the figure).

In order to look into the details of the scattering differences between the probe and the no-probe case, we also performed pair subtraction prior to STFT analysis. An example where the data for the missile with no probe and a short circuit (7USA3) is subtracted from the scattering data for the missile with a probe and an open circuit (6USA9) is shown in Figure 7. Note in the top plot that the strong isolated responses seen in the previous figures have been suppressed.

The initial scattering term (the frustrum scattering) shows a dispersive and strong term shifting from 2.5 GHz to 1.8 GHz over the time frame from 0.2 ns to 2.5 ns. There are also response terms at 4.2 GHz and 6 GHz (also somewhat dispersive). The scattering at 5.2 ns (frustrum to base to probe) extends from 6.5 GHz to DC and is somewhat dispersive also.

The scattering at 9 ns (frustrum-base-frustrum-D-dot probe) is now very extended in frequency. The strongest response is at 2.7 GHz, but there are other terms at 5 and 7 GHz and above. These terms are not dispersive. This plot indicates that the double and triple diffraction mechanisms are quite strong, Their behavior can be extracted using the STFT processing.

An important question is whether the multiple terms in Figure 7 at 9 ns are caused by multiple scattering of

attached modes on the outside of the skin of the missile or if they represent terms that couple to the inside of the missile shape via. the open circuit and probe and travel to the base and back on the inside of the missile. Both cases would imply scattering from the neighborhood of the frustrum/probe to the base and back to the frustrum/probe and then to the D-dot probe. One way of determining the difference is demonstrated in Figure 8. In this figure we subtract data for the case of the probe with the short circuit on the inside (7USA2) from the data for the missile with the probe with an open circuit on the inside (6USA9). The only difference in the two cases is the open and short circuit on the inside of the 5 cm probe on the nose. Note in the figure the strong response at 9 ns. This is the time that corresponds to a signal traveling from the tip to the base and back to the tip where it re-radiates to the D-dot probe. We can speculate that the strong multi-frequency response must be due to internal propagation. Unfortunately, the change from the short to the open circuit also changes the impedance and resonant behavior of the 5 cm probe itself. Thus even though the signal may travel on the outside of the missile, the change in the 5 cm probe impedance would change the behavior of the signal from one data set to the other. That could account for the response we see at 9 ns in Figure 8 also.

Figure 9 is an attempt to resolve this question using the data provided. In this case, we take the difference of two sets of data, one with a 5 cm probe on the tip (7USA2) and one without (7USA3). Both of these missile shapes have a short circuit on the bulkhead on the inside and thus prevent any signal from propagating inside the missile. Note that there is a strong response at 1.6 GHz at 1.5 ns which would correspond to a direct probe effect. On the other hand, there is no strong response at 9 ns, which would be a tip to base to tip to D-dot probe term. This implies that the external signal mode is very small and that the response shown in Figure 8 at 9 ns is due to internal propagating signals.

This is still not a definitive proof of internal propagating modes. To help distinguish internal and external modes, one may put a baffle plate at some distance from the base of the missile shape on the inside. This would change the timing and resonant behavior of the internal modes while leaving the external modes unchanged.

We have also applied the STFT analysis to the integrated data set corresponding to Figure 7. The result is shown in Figure 10. Once again, we are subtracting the missile shape with no probe and an internal short from the case where there is a 5 cm probe on the tip (7USA3) and an open circuit on the inside (6USA9). (Remember that the color scale is now logarithmic.) The response shows similar general behavior as given in Figure 7

6. WIGNER ANALYSIS

6.1. The Wigner Algorithm
The short time Fourier transform algorithm has been used to find the frequency spectral behavior of the scattering from the missile shape as a function of time. This algorithm uses a fixed time window for all frequencies, and thus has a reduction in frequency resolution as the frequency band is lower. Thus a major limitation is in finding an appropriate window for the tradeoff between time and frequency resolution. In many applications, this is undesirable because the requirement is that the spectral resolution must be as high as possible over the signal processing domain.

This problem is addressed directly by considering the energy distribution versus the variables of time and frequency. A non-linear phase space energy distribution approach was originally developed by Wigner [1] in a quantum mechanics context. The theory is developed in several references [1, 2 and 3] and in the Moghaddar Ph.D. dissertation, [24], and will not be re-derived here.

The Wigner distribution, although non-linear, satisfies the condition that a summation over frequency (or time) at a particular time (or frequency) yields the total energy at that time (or frequency). This is called satisfying the marginals.

The most important property of the Wigner distribution is that it can be shown that it has the highest signal concentration in the time-frequency plane. As a serious disadvantage, however, since it is a non-linear technique it generates spurious cross-terms in that plane. This means that interpretation can be very difficult. This is why we have chosen to perform the STFT first and to attempt interpretation of the STFT data first in this study.

The algorithm [following Moghaddar] is given for a discrete-frequency signal S(n) as

$$W(t,n)=\frac{1}{2\pi\cdot(2N+1)}\sum_{k=-2N}^{2N}S(n+k)\cdot S(n-k)\cdot e^{j4\pi ktf\delta l}$$

where:
    n   is the discrete variable
    f   is the frequency
    t   is the continuous variable

In this particular work, we consider the discrete variable as the electromagnetic impulse response of the target. Thus we exchange time and frequency in the above equation and remember that the impulse response is a scalar and not complex.

6.2. The Wigner Algorithm Applied to the Tilted Missile
We can demonstrate the Wigner algorithm using the 45 degree missile orientation data. Examples are shown in

Figure 10 which can be compared with Figure 3 (7USA3; 45 deg; no probe, short inside). The STFT shows the initial response (0.5 ns) to be a single diffuse return at 1.5 GHz or so, while the Wigner (4.5 ns) resolves the return into 3 or 4 individual terms. There is a return 5 ns later in both cases, but the Wigner resolves it into two terms. Finally, at 9 ns later, both algorithms show a term which lowers in frequency in the later (0.5 ns) phase, but the Wigner shows another low frequency term. Both techniques reveal a term near 6 GHz at 9 ns (identified as a delay corresponding to "top to base to D-dot" in section 5.2) from the initial return, but once again the Wigner resolves it into two terms approximately 0.4 ns apart. This 0.4 ns corresponds to a spacing delay on the target (two-way) of 6 cm spacing. This is a reasonable spacing across the 8.5 (tilted) top of the frustrum or it may correspond to a extra delay based on the diagonal extent of the double bounce.

We can also compare the Wigner for 7USA2 (45 deg.; probe; short inside) shown in Figure 11. We expect to see a change in those terms where the probe influences the result. In fact, at this angle, we expect all of the terms to involve the top of the missile and thus can look for effects in each mechanism. A careful comparison shows some differences in the later time of the cluster of terms between 4.5 and 5.6 ns (I.E. the term near 5.6 ns). We also see small terms near 14.5 ns in Figure 4 (with probe) that are not seen in Figure 10 (no probe). These may be caused by the addition of the probe on the tip. Note the higher resolution of the scattering terms. Also note the terms near 7 and 11 and 17 ns. These are due to interactions that are the result of the non-linear nature of the Wigner algorithm. (There are ways to reduce this effect, but they are not necessary here.)

We also applied the Wigner algorithm to 6USA9 (45 deg; probe; open inside) as seen in Figure 12. In this case, the only change between this figure and Figure 4 is the change on the inside of the bulkhead connector from a short circuit to an open circuit. There are some changes near 5.9 ns (the later part of the initial term, tentatively identified as a scattering from the frustrum area directly to the d-dot probe. These changes are not specific enough to be unambiguously identified as caused by the change from the short to the open circuit.

Once again, it is seen that the resolution in the time-frequency domain is better for the Wigner than for the STFT. The identification of specific terms is still difficult because the experimental set-up was not specifically for this type of analysis.

6.3. The Wigner Algorithm; Summary
We have shown the use of the Wigner time-frequency algorithm to evaluate the impulse scattering from the missile shape. The algorithm has been shown to have higher resolution than the STFT. The interaction terms were not a problem in this application, mostly because

the STFT resolution was sufficient to resolve most of the interaction ambiguities. It is never-the-less possible that some of the fine-grain behavior that we discussed was actually some type of interaction.

The experiments could be modified in the future to specifically reveal some of the specific individual mechanisms in the interaction. Very careful (internal motor driven) probe extension as data are taken is one example.

7. WAVELET ANALYSIS

7.1. Wavelet Analysis

It is well-known that it is impossible to perform spectral analysis with infinite resolution in the time and frequency domain simultaneously. (We might have a deterministic formula for the time domain behavior, for example, and then we could get infinite resolution in the frequency domain. This is what happens with model-based superresolution techniques.) The most common spectral analysis technique (the Fourier technique) is actually a type of convolution between finite data sets and sinusoids. (Windowing techniques make up for some of the flaws of this approach.) The limits of the resolution are fundamentally due to the finite nature of the data sets and the convolution process.

Wavelet spectral analysis recognizes immediately that the data set to be processed is to be convolved with a specific waveform (the wavelet), and specifies a resolution in frequency and in time by the careful design of the wavelet [13, 14, 16, 20, 21, 22 and 24]. The result is a time-frequency domain representation of the waveform where the resolution is optimized over the domain of time and frequency. Entire books have been written on the wavelet transform, and thus a full discussion is beyond the scope of this paper. See the very nice survey article by Rioul and Vetterli in the IEEE Signal Processing Magazine [13], or try the book *A Friendly Guide to Wavelets* by G. Kaiser [21]. There are also a number of world wide web pages on the Internet where more background can be found.

Consider the short time Fourier transform (STFT) discussed earlier in this report. In this case, a sinusoid is windowed to produce a finite time duration sinusoid which is then convolved with the waveform. Actual implementation may change the order of the process, but we must have a triple product of the original waveform to be evaluated, the window and the sinusoid. In the STFT, note that the window size is fixed as a function of frequency and time. Thus as the sinusoid is scanned over frequency, the number of cycles in the window varies. Thus the resolution of the transform varies as a function of frequency. One way to correct for this (if we want to) is by creating a window which always contains the same number of cycles. This means that we can simply modify the STFT so that the window is a function of

frequency and then perform the convolution. The result is a very useful type of wavelet. In fact, the STFT and the wavelet transform can be generalized as simply two types of cross-ambiguity functions and they have the same general properties [13].

Before we get into the application of the wavelet transform here, the important concept of "scale" must be discussed. The development of the wavelet transform leads to issues of orthogonality, completeness and energy conservation (of course). The result is that the natural units to use instead of frequency is "scale." This is because equal increments of frequency would produce variations in the resolution and thus we would move from an optimal frequency domain sampling to over-sampling or under-sampling. The only scheme which works is a uniform "scale" axis. (Refer to the references.) Thus the plots given here will be given in units of "scale" versus time.

We will use the discrete wavelet transform (DWT) here. Simply put, we use a prototype waveform (a "mother wavelet") and convolve it with the d-dot probe time domain waveform. All other wavelets are stretched or compressed versions of the prototype. The wavelet prototype $h_a(t)$ is given as

$$h_a(t) = \frac{1}{\sqrt{|a|}} \cdot h(\frac{t}{a})$$

where a is the "scale factor." Thus the definition of the DWT is

$$DWT(t,a) = \frac{1}{\sqrt{|a|}} \sum (x(t) \cdot h(\frac{t-t}{a}) \cdot \Delta t)$$

Note that variations in time and scale are imbedded in this convolution. Also note the inefficiency of this calculation because the tricks used in the "fast" Fourier transformation are not used. (A "fast" DWT is available, however; see [21])

With this version of the wavelet transform, we can form a specific relationship between frequency and scale. In this particular case, it is

$$\text{scale} = (f/F)^{(0.25)}$$

where F is the folding frequency and $F = 1/\delta$, where δ is the sampling time increment. In this case, $\delta = 0.02$ ns, and thus $F = +/- 25$ GHz.

An example data file was created to demonstrate the behavior of the wavelet transform. The test waveform is the sum of the following four signals.

1. An impulse (a single data point) of amplitude three at 1.1 ns.
2. A sinusoid of amplitude 1 and frequency 2.3 GHz starting at 4.3 ns.
3. A sinusoid of amplitude 1 and frequency 6.7 GHz starting at 9.7 ns.
4. A sinusoid of amplitude 1 and frequency 12.4 GHz starting at 12.4 ns.

If we evaluate this waveform using the STFT, we get the result shown in Figure 13. Note that the 50 data point span is such that the start time of the STFT response is at 0.5 ns. This is early enough to capture the single point event, but there is no effect that can be seen. The first waveform begins at 4.3 ns and has frequency 2.3 GHz, and the second waveform begins at 9.7 ns and has frequency 6.7 GHz as expected. The response images for these waveforms are very clear. The third waveform is above the top of the plot. This is because we stopped plotting the response at 9 GHz because the impulse radar has an upper frequency limit of approximately 8 GHz, and we wish to see the details of the response below 8 GHz.

The wavelet analysis of this waveform is shown in Figure 14. In this figure, the impulse can be seen in the upper plot. We see the wavelet response of the impulse at 1.1 ns, where the STFT was not able to resolve it. Note that most of the energy is at the high frequency limit of the analysis, as we expect for the very short time impulse modeled here. Next, we see the three sinusoids starting at 4.3, 9.7 and 12.4 ns. They have scale factors of 0.55 (2.3 GHz) , 0.72 (6.7 GHz) and 0.8 4 (12.4 GHz). Note that all three waveforms are shown. In fact, the scale from 0 to 1 is the entire frequency span of (0 to 1/2t, where t is the time increment).

7.2. Wavelet Analysis Applied to the Tilted Missile
The application of the Continuous Wavelet Transform (CWT) to the 45 degree orientation data is shown in Figures 15 and 16. The configuration is summarized below.

- Figure 15-   7USA3; no probe internal short
- not shown here- 7USA8; no probe; internal open
- not shown here- 7USA2; probe; internal short
- Figure 16-   6USA9; probe; internal open

Careful observation of these plots does not show any obvious change as the probe is installed or removed. If we make a careful comparison between the STFT results and the CWT results, we note that the initial response (2.2 ns at 0.5 scale = 1.56 GHz) and the second response (11.5 ns and 0.5 scale = 1.56 GHz also) are very similar in relative time and frequency (1.5 GHz corresponds to 0.5 scale = 1.56 GHz).

We subtracted pairs of data from the above data sets and formed the Wavelet transform on the difference waveforms. The plots are given in Figures 17 and 18, and listed below.

Figure 17: 6USA9 - 7USA2; probe on both; internal; open minus shorted

Figure 18: 6USA9 - 7USA3; (probe; open) minus (no probe; short)

In the subtracted pairs, clear multiple interaction effects are seen. Both show clear effects at the time delay corresponding to an internal bounce off the inside bottom (10 to 11 ns.; 0.75 scale). If both data sets have a probe (Fig. 17), no change is seen at 2 ns where the probe scattering effects are expected. If only one has a probe, a clear dispersive probe effect is seen at this location. As was shown in the subtracted STFT data, there appears to be an internal reflection mechanism, where the signal penetrates into the missile body through the probe and the open ended coaxial bulkhead connector on the inside.

We especially note that the response at a delay of 6.5 ns after the initial response is probably "real." It is ambiguous in the STFT results, but seems clear and specific in the wavelet images. Unfortunately, if we look at the earlier 45 degree write-up, we see no specific mechanism that corresponds to this type of time delay. This response mechanism needs more study.

7.3. Wavelet Analysis; Summary
The wavelet analysis has shown very close agreement with the STFT results in this case. The mechanisms involved in the mock missile do not have a form which would be critical to the differences in the resolution abilities between the STFT and the CWT. If there were more dispersive mechanisms involved in the scattering from the missile shape, then the differences would become important because the curves would have specific dispersive signature shapes [12, 14, 15, 16, 18 and 19].

We have shown the general application of this type of analysis to impulse measurements such as can be obtained from the time domain range and the d-dot probe. Other types of scattering mechanisms, such as propagation through dispersive media (plasma or earth) or scattering from resonant targets (cavities or ducts) would more critically depend on wavelet analysis.

8. BISPECTRAL ANALYSIS

8.1. Bispectral Analysis
So far, we have been discussing transformations from the time domain to the time-frequency domain. The goal

has been to identify the spectral response of various scattering mechanisms. The direct scattering mechanisms are relatively easy to characterize because there is a direct relationship between the time of occurrence and the range domain (or down-range) location of the scatterer. Many of the terms in the response profile, however, are the result of multiple interactions. These terms are difficult to specifically identify because they appear no differently in the down range profile than the direct scattering terms. There is a difference, however. The interactive terms are related in a specific way to the two direct terms involving the interacting scatterers. An algorithm that can reveal these relationships is the bispectrum.

The bispectrum is defined as the two-dimensional Fourier transform of the third moment sequence or the bicorrelation function $R(t_1,t_2)$ [4 to 8]. If we have scattering measurements in the time domain, h(t), the bicorrelation function is defined as

$$R_h(t_1,t_2) = \int_{t=-\infty}^{t=\infty} h^*(t)\, h(t+t_1)\, h(t+t_2)\, dt$$

(remember that h(t) is scalar) and thus the bispectrum of h(t) is defined as

$$B(w_1,w_2) = \int_{t_1=-\infty}^{t_1=\infty} \int_{t_2=-\infty}^{t_2=\infty} R_x(t_1,t_2)\exp(-j(w_1 t_1 + w_2 t_2))dt_1\, dt_2.$$

Note that if the data are corrupted by noise, the bispectrum (which represents an integration over the domain of t_1 and t_2 tends to the bispectrum of the noiseless data. In other words, the triple correlation of the noise approaches zero [5 and 8].

An alternate formulation of the bispectrum is to take the triple product of the spectrum that results from the Fourier transform of the impulse response h(t). Thus if

$$S(w) = \int_{t=-\infty}^{t=\infty} h(t)\exp(-jwt)\, dt,$$

then it can be shown that

$$B(w_1,w_2) = S(w_1)S(w_2)\, S^*(w_1+w_2).$$

This is the generalized form of the bispectrum. In this form, it permits relationships between spectral responses at two different frequencies to be discovered. Our problem is different, however. Our problem is to discover relationships between two different responses in the time domain. We will make this change below.

It has been shown for radar scattering that it is possible to exchange the time and frequency parameters in the above argument and arrive at a "bi-time" response or, since time is related to distance by the speed of light, the "bi-range profile" of a radar target [8 to 11]. Using a similar argument (with t and ω interchanged), we arrive at a form of the bi-time response as

$$B(t_1,t_2) = h(t_1)\, h(t_2)\, h(t_1+t_2).$$

Where h(t) is the (band limited) radar target impulse response. What we are saying is that two response terms in the impulse response at t_1 and t_2 can be shown to be due to an interaction that produces a third term at t_1+t_2 if there is a response at the product of $h(t_1)\, h(t_2)$ and $h(t_1+t_2)$. This information has been used for radar target identification [5, 6, 8 and 9].

The above formula can be shown to only apply to a case where the multiple interaction is along the axis of propagation of a monostatic radar beam. It also has a problem with the zero time offset [10]. In the next section, we will modify the algorithm to apply specifically to our 45 degree missile.

8.2. Bispectral Analysis for the 45 Deg. Case

As we can see, the bispectrum can be used to analyze radar direct scattering and interactions when the multiple scattering is along the same axis as the incident beam. In our case, we have a missile shape tilted at a 45 degree angle and the d-dot probe is in the near field and is at a distinct bistatic angle with respect to the incident signal. The bispectral algorithm must be modified to fit this specific configuration in order to evaluate multiple interactions. Consider the geometry of a 45 degree tilted object as shown in Figure 19.

In this figure, (1) an incident signal arrives at points A and B and scatters bistatically to the D-dot probe, (2) the incident signal arrives at point A, scatters to point B and then back to point A, and then to the D-dot probe. We wish to use the bispectral concept to evaluate the interaction between points A and B. Using this geometry, (and the known value of the intercept point, x) for every pair of points in time t(A) and t(B), corresponding to the reception of a signal from generalized points A and B, we can compute the time of the interaction term t(C). We then form the triple product

$$B(t(A),t(B)) = t(A) * t(B) * t(C).$$

The two dimensional image, B(t(A),t(B)), is then generated. The magnitude of the response at each point in the image is proportional to the correlation of the product t(A)*t(B) with t(C). This magnitude will be non-zero when there is an interaction term between A and B.

Using (relatively) simple trigonometry, we have derived the generalized value of t(C) for given values of t(A) and t(B) and the intercept point x.

8.3. Bispectral Analysis Applied to the Tilted Missile Data

We applied the 45 deg. modified bispectral analysis to the 45 degree missile impulse scattering data. The 45 deg. modified bispectral image for file 7USA3 (no probe; internal short) is shown in Figure 20. Ignoring the data on the main diagonal (t(A) = t(B) = t(C)), we can see several specific interaction terms. They are listed below. All timing/mechanism estimates are based on geometrical considerations.

1. Vertical index 65; horizontal 110 (remember that each index is 0.02 ns)
This timing corresponds to an interaction between a reflection from the top of the frustrum and the bottom of the frustrum. (Top of frustrum to bottom of frustrum to d-dot).

2. Vertical index 55; horizontal index 130
This timing corresponds to an interaction between a reflection from the top of the frustrum and something with a delay time 1.4 ns longer. There is no obvious mechanism except perhaps an attached mechanism that forms a creeping wave around the back of the missile cylinder/frustrum and returns to the D-dot probe. More study of this is called for. It may be a false correlation simply because of an overlap with other terms.

3. Vertical index 65; horizontal index 180
This timing corresponds to an interaction between a reflection from the top of the frustrum and a point 3/4 of the way down the cylinder part. There does not seem to be a scatterer at this point, but perhaps we are actually seeing an interaction between the nose of the missile and the base of the missile. More detailed study of this algorithm is called for.

4. Vertical index 120; horizontal index 180
This timing corresponds to an interaction between a direct creeping wave at the bistatic 45 degree specular point and a term 2.4 ns later. No mechanism is seen there. It may be an accidental overlap of other terms.

5. Vertical index 180; horizontal index 210
This timing corresponds to an interaction between the near side of the base and the far side of the base of the missile shape.

We also applied this algorithm to the data sets from 6USA8 (no probe; internal open), Figure 21, 7USA2 (probe, internal short) - Figure 22, and 6USA9 (probe, internal open) - Figure 23.

If we compare these figures, we see a change as we go from the no-probe case to the probe case at index 55 vertical 180 horizontal. This timing might correspond to an interaction between the probe and the bottom of the missile shape (as discussed in item 3 above). This is speculation.

If we compare the internal open with the internal short case, we also see a possible effect at 55 vertical and 180 horizontal (the same location as described just above for the probe versus no-probe case). Once again, we can speculate that there is some effect involving the change caused by the open versus short case which effects the probe versus bottom of missile interaction.

8.4. Bispectral Analysis; Summary

In summary, what we see in the image seems to show multiple interaction terms between the top and bottom of the frustrum and between the near side and the far side of the base. We also may see an interaction between the frustrum and the base of the missile shape.

There also seem to be some "accidental" bright spots. These could be caused by the accidental correlation of some strong terms in the time domain response. Note, for example, the off-diagonal lines in the image. They originate at specific bright points and cause "bright spots" where they cross the horizontal or vertical response lines of other strong scattering mechanisms. These intersection spots are likely not true interaction terms.

Clearly, this modified bispectral interaction analysis shows promise. We may actually be able to specifically identify interaction terms in the impulse response. More research on this modified application of the bispectrum is called for.

9. CONCLUSIONS

9.1 Time-Frequency Representations for Impulse Scattering Analysis

We have shown applications of the Short Time Fourier Transform, the Wigner algorithm, the wavelet transform and the bispectrum to the problem of analysis of the resonant behavior of a generic missile shape. The missile was tilted 45 degrees toward the impulse radar.

All three analysis techniques were able to show

1) the direct reflection mechanisms and their frequency spectrum

2) multiple reflection mechanisms involving
    (a) multiple reflections (interactions) from the nose region to the rear
    (b) multiple reflections (interactions) across the nose region and the rear region

Note that any multiple interaction mechanism is, in fact, a type of resonance.

A particularly effective technique was to subtract two data files involving only a specific change in the radar target (with and without a probe, for example). This technique was able to extract specific interactions that could not be seen directly.

Each analysis technique had its own area of specific application.

(1) STFT; This technique is linear, robust and fast. In the target under test here, the reflection mechanisms were far enough apart to permit the STFT to specifically identify most scattering terms.

(2) The Wigner algorithm; This technique has more resolution in both the time and frequency domain than the STFT, but the non-linear nature of the transform generates "mixing" terms in the image that can be confusing. Fortunately, the STFT was able to help distinguish the false terms from the real terms.

(3) The wavelet transform; This technique permits optimal resolution in the time and frequency domains simultaneously so that both short time events and narrow frequency events can be revealed in a single image. The logarithmic frequency index (the "scale" index) permits the entire frequency span to be seen. Resolution is optimized to this scale index, rather than to frequency.

(4) The bispectrum; The bispectrum is specifically appropriate to reveal multiple interactions. With the earlier algorithms, it was necessary to identify multiple interactions by their location in the time domain (ambiguous at best). The bispectrum permits direct calculation of the correlation of the interaction terms. In this specific 45 degree missile shape example, a modified version of the bispectrum was used. The result reveals several interaction terms. Interpretation remains difficult because of the "accidental" correlations which occur.

9.2 Evaluation of the Missile Shape

We see direct scattering (at 1 GHz) from specular and diffraction sides and edges of the missile. The frustrum is a particularly strong scatterer when the missile is tilted toward the radar d-dot probe.

We see interaction terms between the frustrum neighborhood and the rear of the missile, as well as what appears to be a resonance involving the length of the body (1 wave-length). Note that a resonance and a multiple interaction are identical effects in most cases.

We also may be able to see effects due to the internal resonances of the missile (differences between internal

"short" and internal "open"). These effects occur when the probe on the nose of the missile conducts energy directly into the interior of the missile. They are easy to confuse with external attached mode multiple interaction with the frustum area.

10. REFERENCES:

1. Wigner, P. E. "On the Quantum Correction for Thermodynamic Equilibrium," *Phys. Rev.*, Vol. 40, pp. 749-759, 1932.
2. Wigner, P. E. "Quantum-mechanical Distribution Functions Revisited," in *Perspectives in Quantum Theory* (W. Yougrau and E. A. van der Merwe, eds.), Cambridge, MA: M.I.T. Press, 1971.
3. Classen, T.A.C.M. and W. F. G. Mecklenbrauker, "The Wigner Distribution - A Tool for Time-Frequency signal Analysis; Part 1: Continuous-time Signals," *Philips J. Res.*, Vol. 35, no.3, pp. 217-250, 1980.
4. Walton, E. K. and I. Jouny, "Application of Bispectral Techniques to Radar Scattering Signatures," Ant. Meas. Tech. Asso. Meeting, Monterey, CA, October 1989.
5. Walton, E. K. and I. Jouny, "Bispectrum of Radar Signatures and Application to Target Classification," Radio Science, Vol. 25, No. 2, pp. 101-113, March 1990.
6. Walton, E. K. and I. I. Jouny, "Target Identification using Bispectral Analysis of UWB Radar Data," First Los Alamos Sympo. on UWB Radar, Los Alamos, March 1990.
7. Jouny, I. I. and E. K. Walton, "Bispectral Analysis of Radar Signals," Joint MTT, IEEE AP-S and URSI Sympo., May 1990.
8. Walton, E. K. and I. I. Jouny, "Target Identification using Bispectral Analysis of UWB Radar Data," In *Ultra-Wideband Radar: Proceedings of the First Los Alamos Symposium,* Ed. Bruce Noel, CRC Press, Boca Raton, 1991.
9. Jouny, I. I., E. K. Walton, F. D. Garber and R. L. Moses, "Applications of the Bispectrum in Radar Signature Analysis and Target Identification," Proc. of the 1991 Automatic Object Recognition Session, Society of Photo-Optical Instrumentation Engineers Meeting, April 1991.
10. Walton, E. K. and I. I. Jouny, "Target Identification in Non-Interactive Clutter using the Bispectrum," Joint URSI Meeting and Int. IEEE AP-S Sympo., London, Ontario, July 1991.
11. Walton, E. K. and I. I. Jouny, "Applications of the Bispectrum in Radar Signatures Analysis and Target Identification," Int. Signal Processing Workshop on Higher Order Statistics, July 1991.
12. Walton, E. K. and Moghaddar, "Time-Frequency Distribution Analysis of dispersive Targets," Workshop on High Frequency Electromagnetic

Modeling of Jet Engine Cavities, sponsored by Wright Patterson AFB, August 1991.

13. Rioul, O. and M. Vetterli, "Wavelets and Signal Processing," IEEE Sig. Proc. Magazine, Vol. 8, No. 4, pp. 14-37, Oct. 1991.

14. Walton, E. K. and A. Moghaddar, "Time-Frequency Distribution Analysis of Waveguide Modes," IEEE AP-S Sympo., Chicago, July 1992.

15. Walton, E. K. and A. Moghaddar, "Analysis of Inlets and Ducts using Time Frequency Distribution Analysis of UWB Radar Signals," Inter. confer. on UWB Short-Pulse Electromagnetics, Weber Res. Instit. of Polytechnic Univ. and IEEE AP-S and MTT Societies, Brooklyn, N.Y., October 1992.

16. Walton, E. K. and A. Moghaddar, "Time-Frequency Distribution Analysis of Frequency Dispersive Targets," in *Ultra-Wideband, Short Pulse Electromagnetics* ed. H. Bertoni et al., Plenum Press, 1993.

17. Cloude, S. R. et al, "Analysis of Time domain Ultrawideband Radar Signals," in *Ultra-Wideband, Short Pulse Electromagnetics* ed. H. Bertoni et al., Plenum Press, 1993.

18. Walton, E. K. and a. Moghaddar, "Time-Frequency Distribution Analysis of Scattering from Waveguide Cavities," IEEE Trans. on Ant. and Prop., Vol. 41, No. 5, pp. 677-680, May 1993.

19. Walton, E. K., A. Moghaddar and Y. Ogawa, "Superresolution Analysis of Frequency Dispersive Scattering," Antenna Measurement Techniques Assn. Meeting, October 1993.

20. Moghaddar A., E. K. Walton and W. D. Burnside, "Time-Frequency Distribution Analysis of Frequency Dispersive Scattering Using the Wavelet Transform," Antenna Measurement Techniques Assn. Meeting, October 1993.

21. Kaiser, G. *A Friendly Guide to Wavelets*, Birkhauser Press, Boston, 1994.

22. Moghaddar, A. and E. K. Walton, "A Data-Adaptive Time-Frequency Representation Applied to the Scattering from a Jet Aircraft," IEEE AP-S Int. Sympo, Seattle, WA, June 19, 1994.

23. Moghaddar, A., Y. Ogawa and E. K. Walton, "Estimating the Time-Delay and Frequency-Decay Parameter of Scattering Components using a Modified MUSIC Algorithm," IEEE Trans. on Antennas and Prop., Vol. 42, No. 10, pp. 1412-1419, October 1994.

24. Moghaddar, A. "A Time-Frequency Representation of Frequency Dispersive Waveguide and Cavity Scatterers," Ph. D. Dissertation, The Ohio State University, December 1994.

25. Walton, E.K. and I. Jouny, "Target Identification Using Bispectral Analysis of UWB Radar Data," in Ultra-Wideband Radar: Proceedings of the First Los Alamos Symposium, Ed: Bruce Noel, CRC Press, Boca Raton, July 1991.
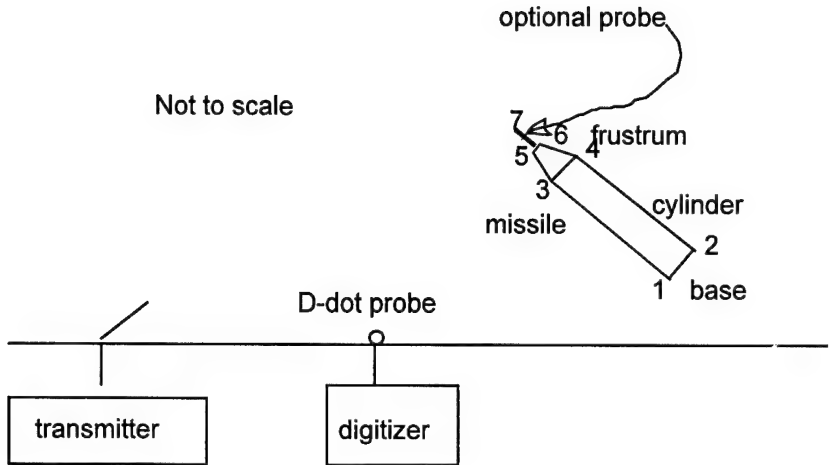
Figure 1. Experimental setup for impulse testing.
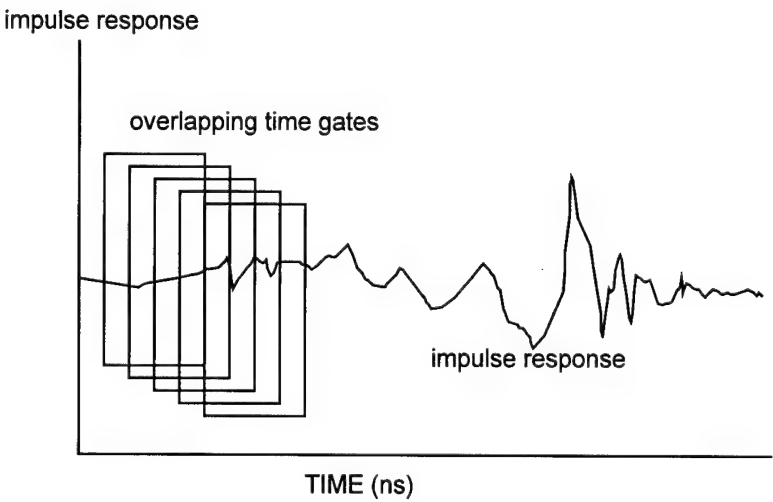
Figure 2. STFT Algorithm sketch.
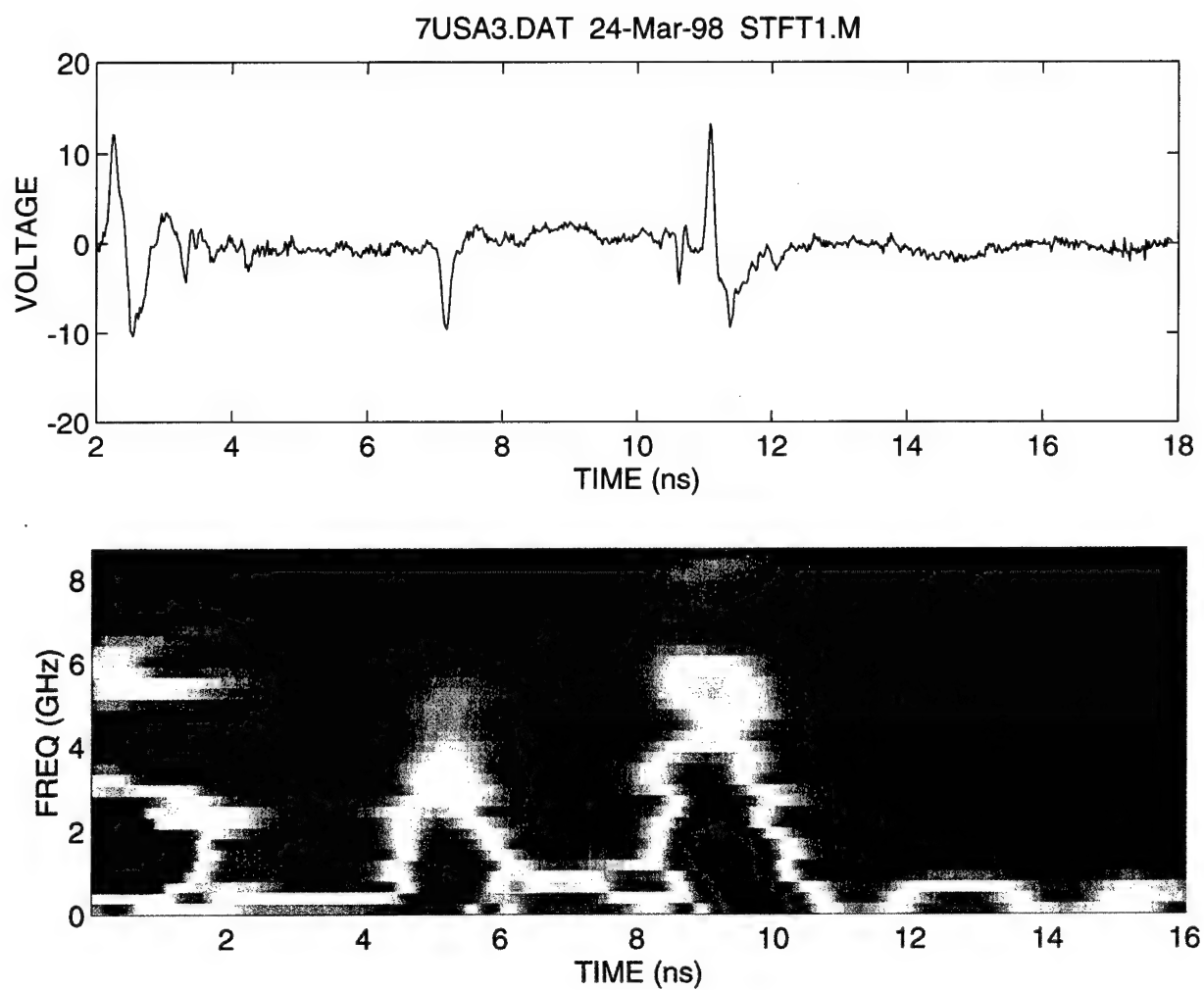
Figure 3. Raw data and STFT image for 7USA3 (45°).
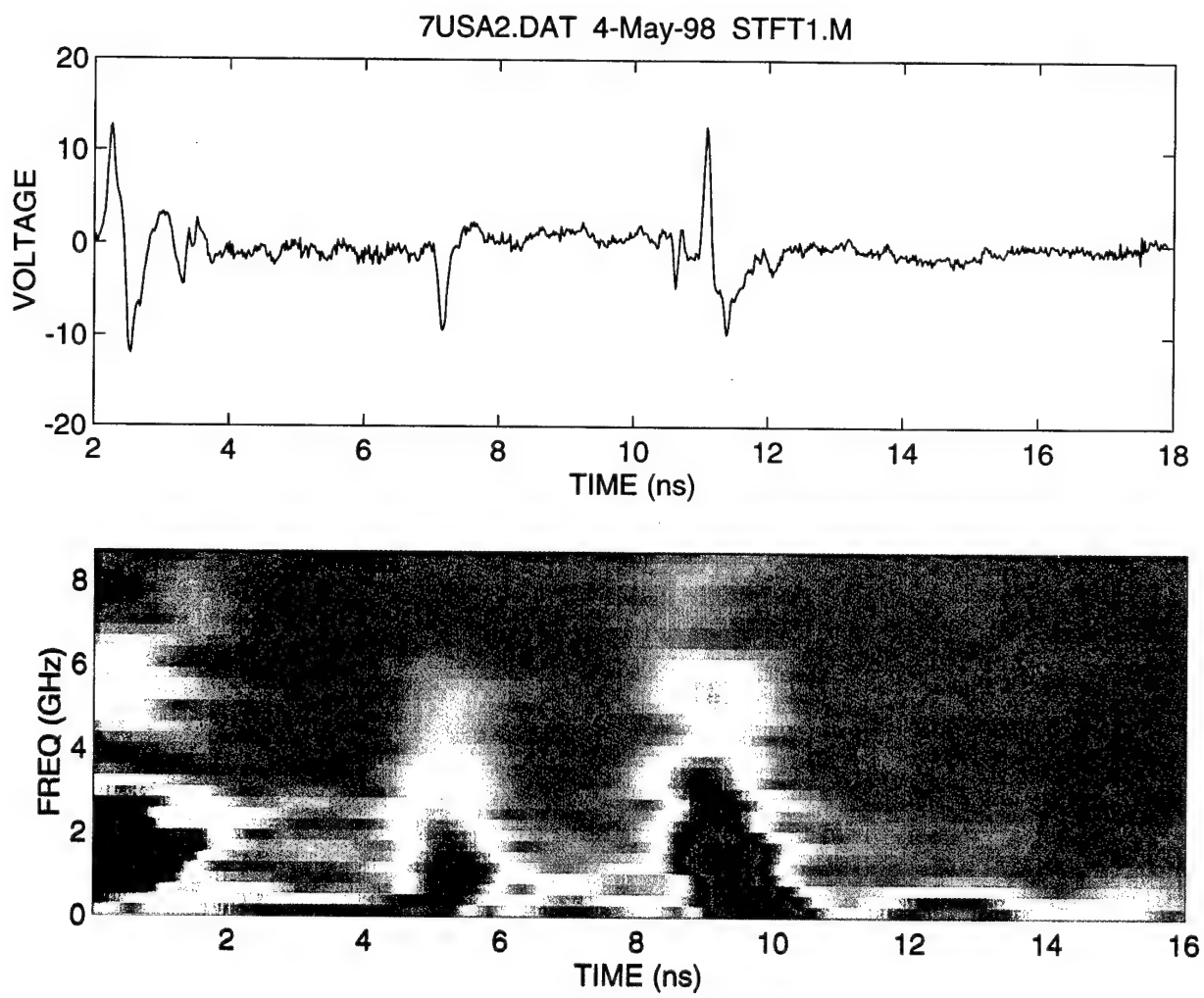
Figure 4. Raw data and STFT image for 7USA2 (45°).
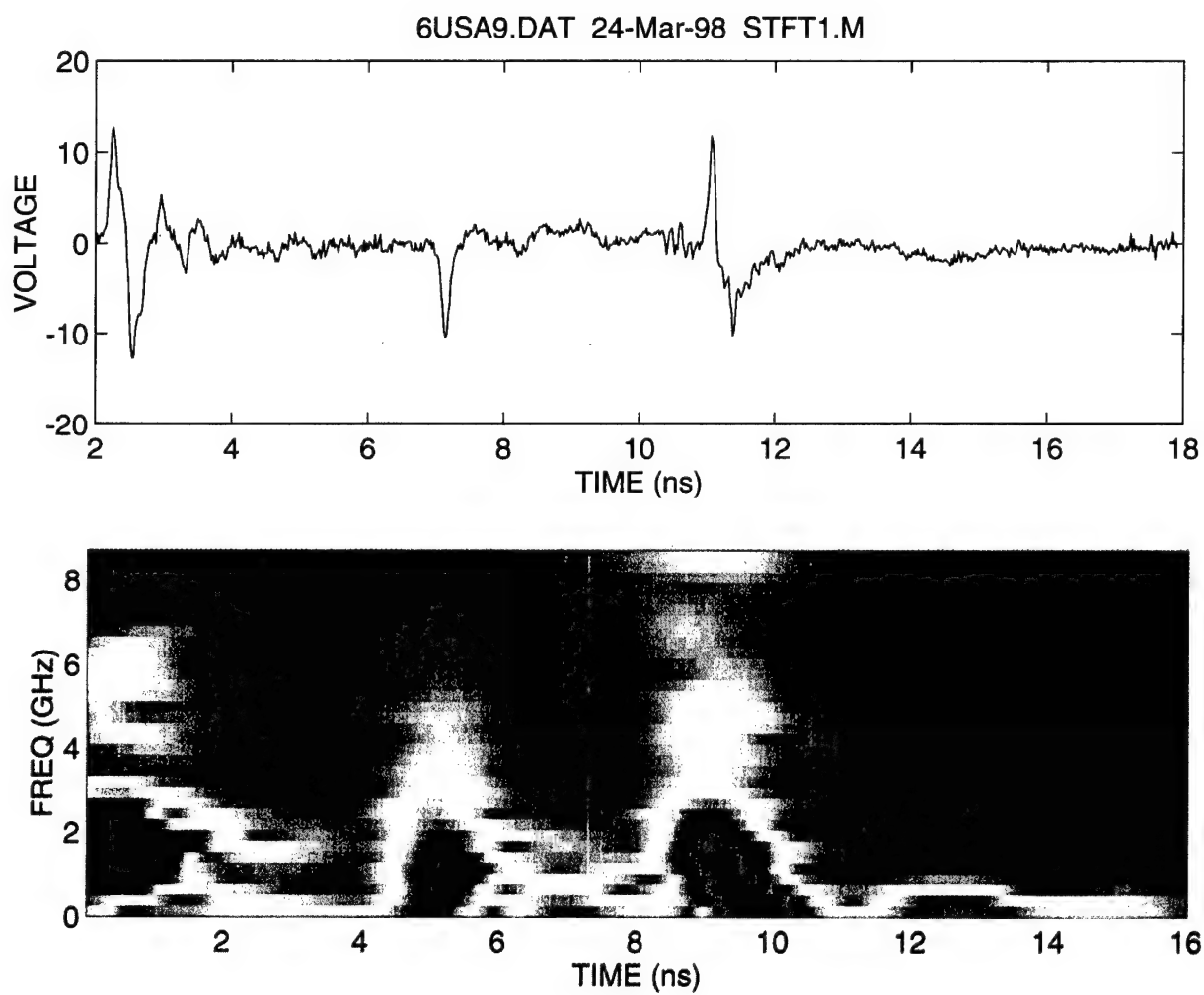
Figure 5. Raw data and STFT image for 6USA9 (45°).
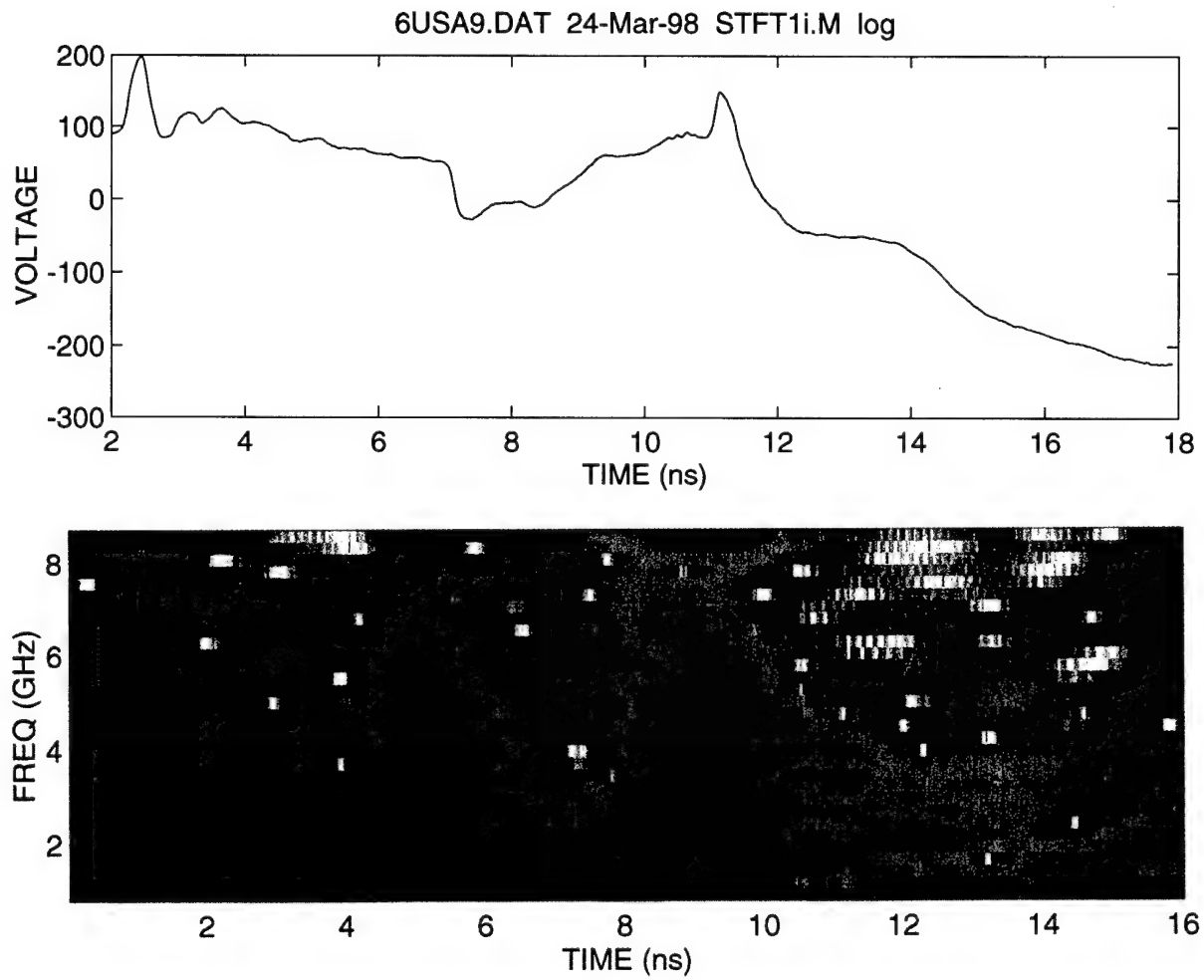
Figure 6. Raw data and STFT image for 6USA9 (45°; after integration).
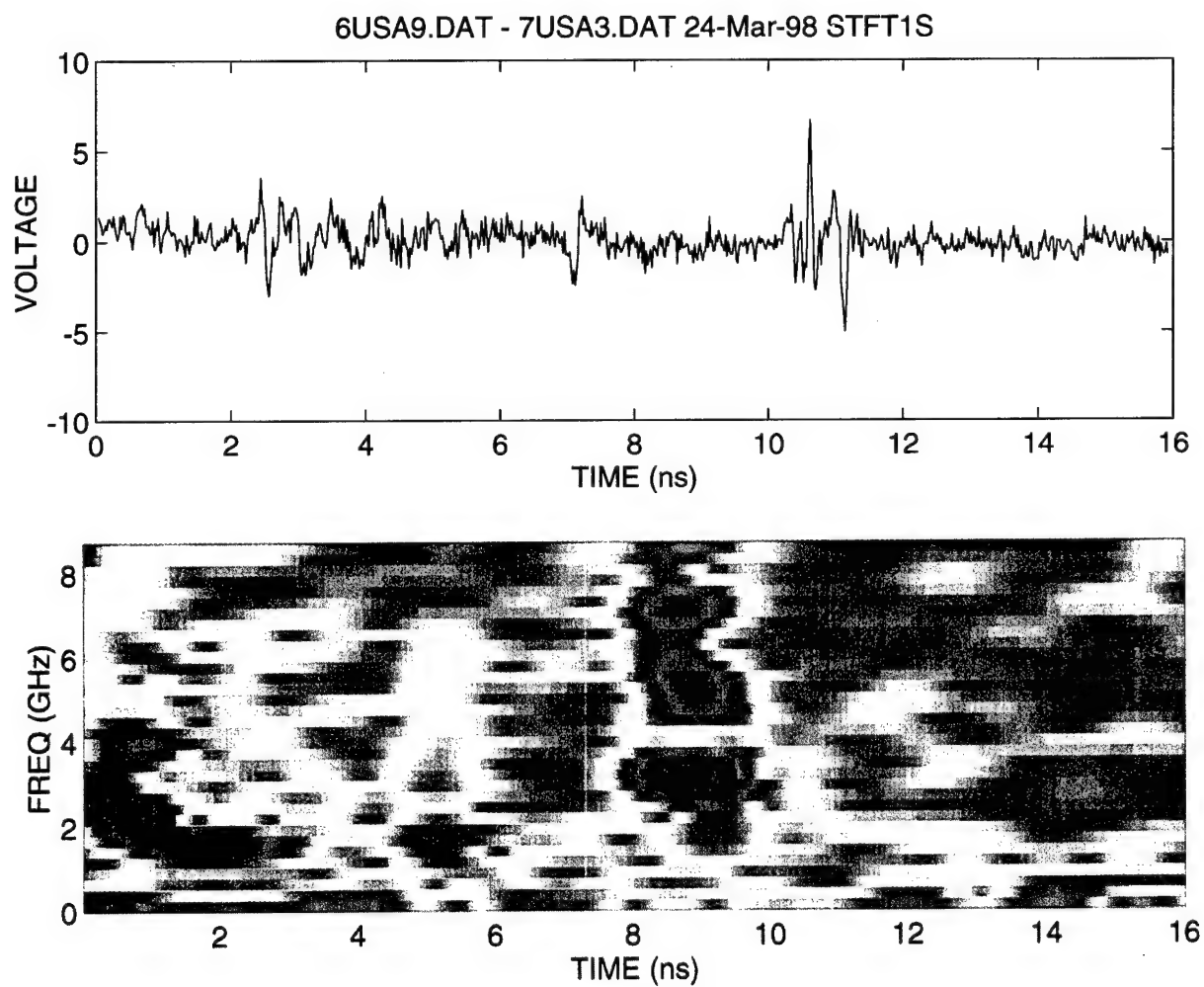
Figure 7. Raw data and STFT image for 6USA9 minus 7USA3 (45°).
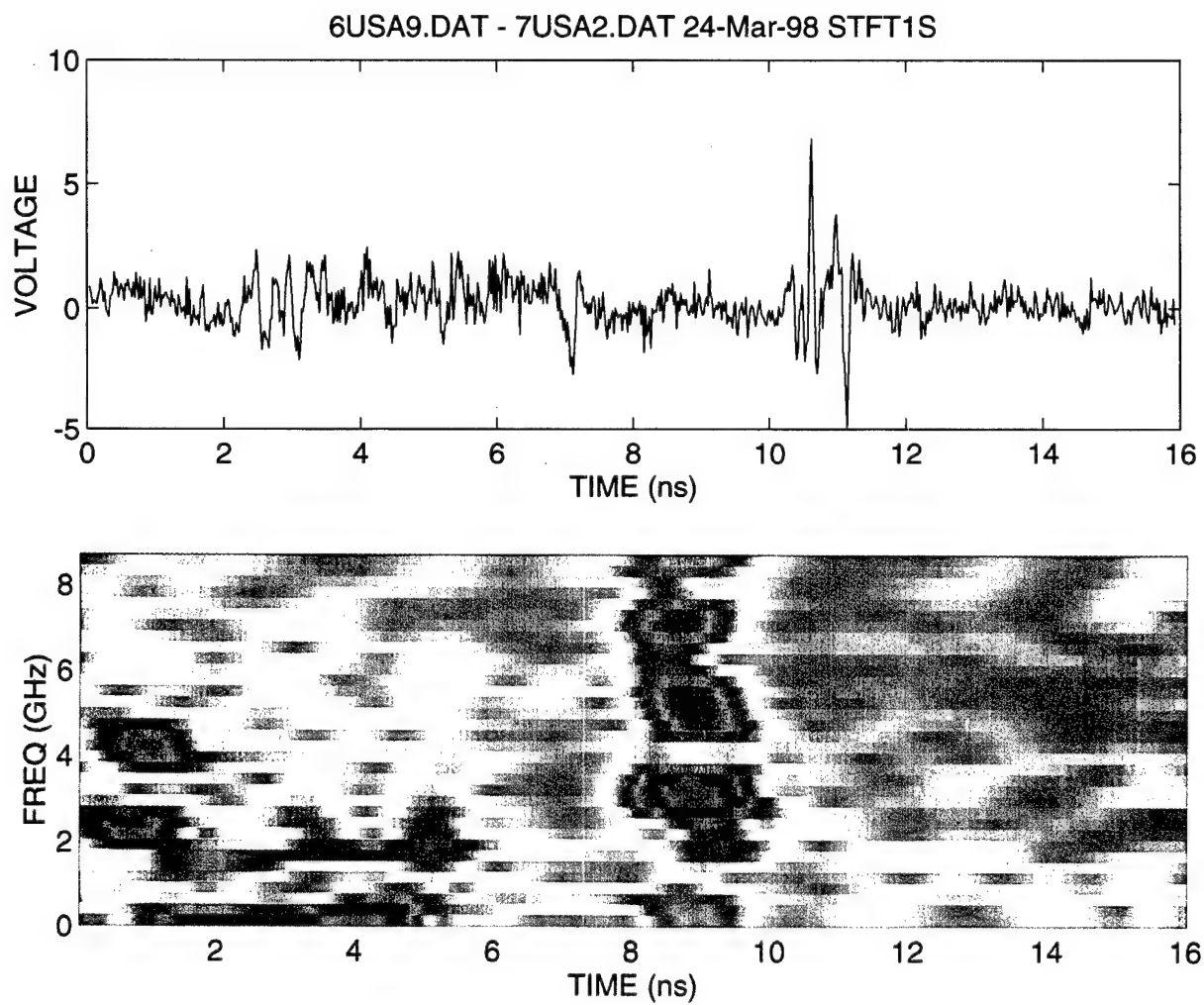
Figure 8. Raw data and STFT image for 6USA9 minus 7USA2 (45°).
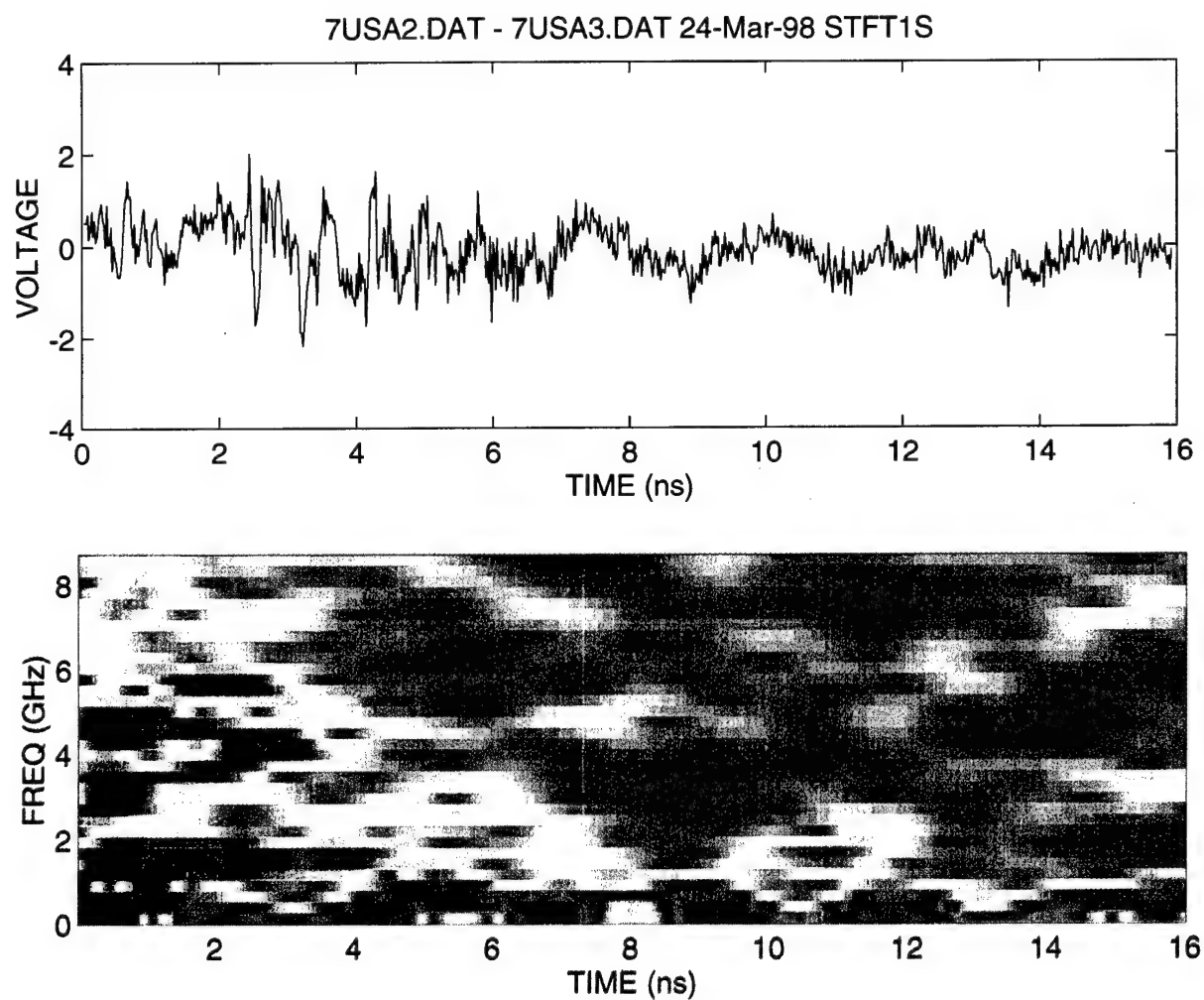
Figure 9. Raw data and STFT image for 7USA2 minus 7USA3 (45°).
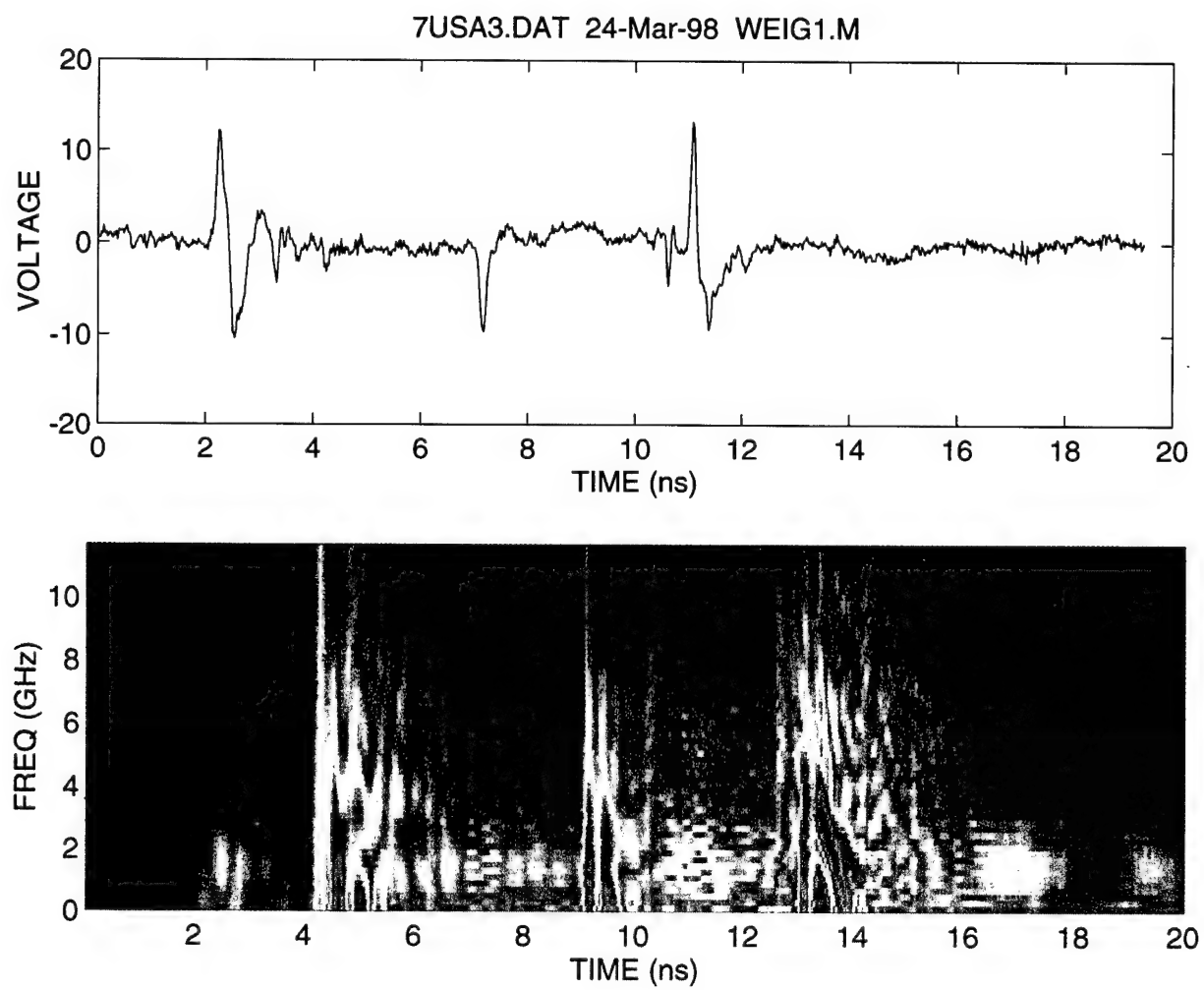
Figure 10. Raw data and Wigner algorithm for 7USA3 (45°).
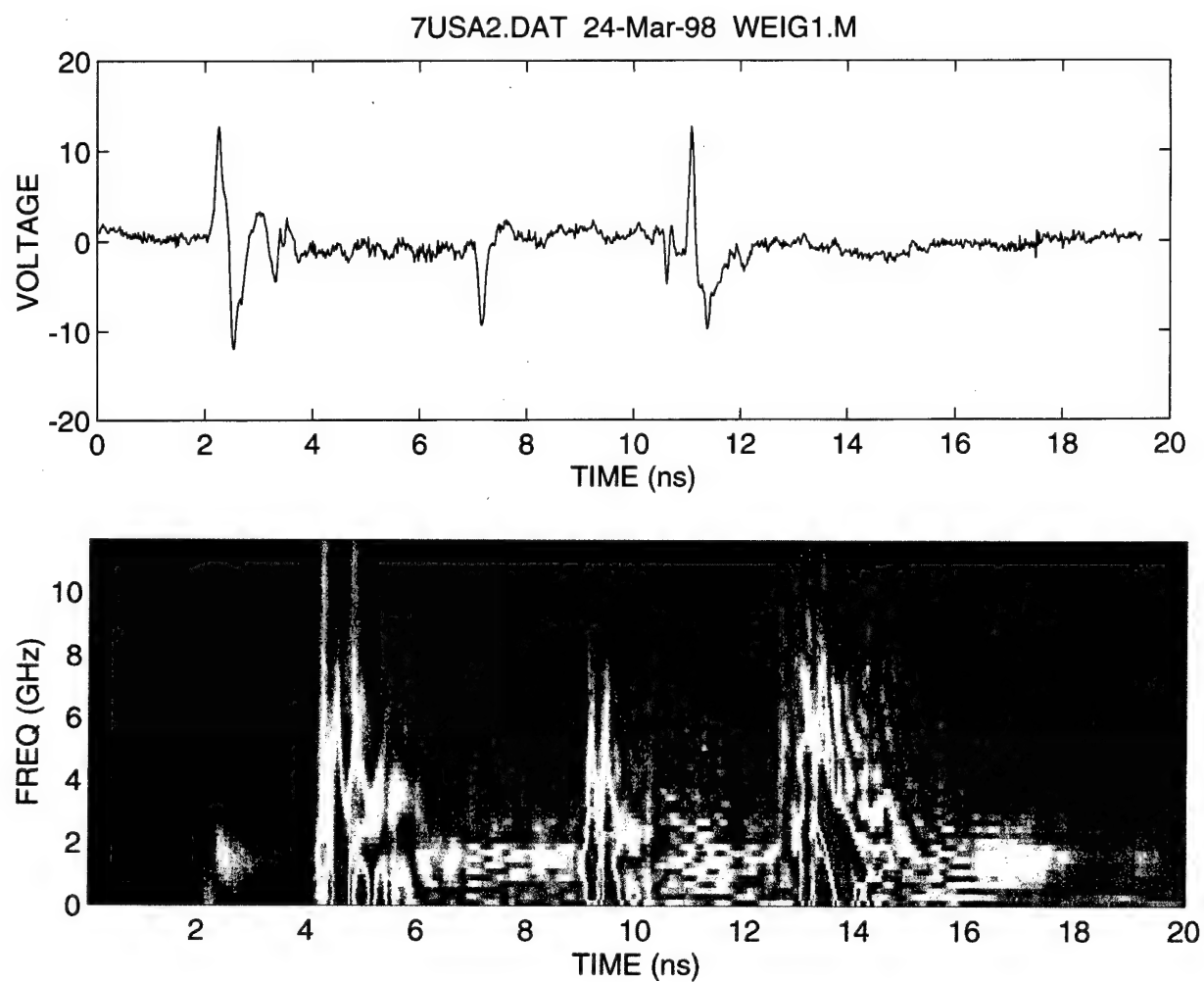
Figure 11. Raw data and Wigner algorithm for 7USA2 (45°).

6USA9.DAT 24-Mar-98 WEIG1.M
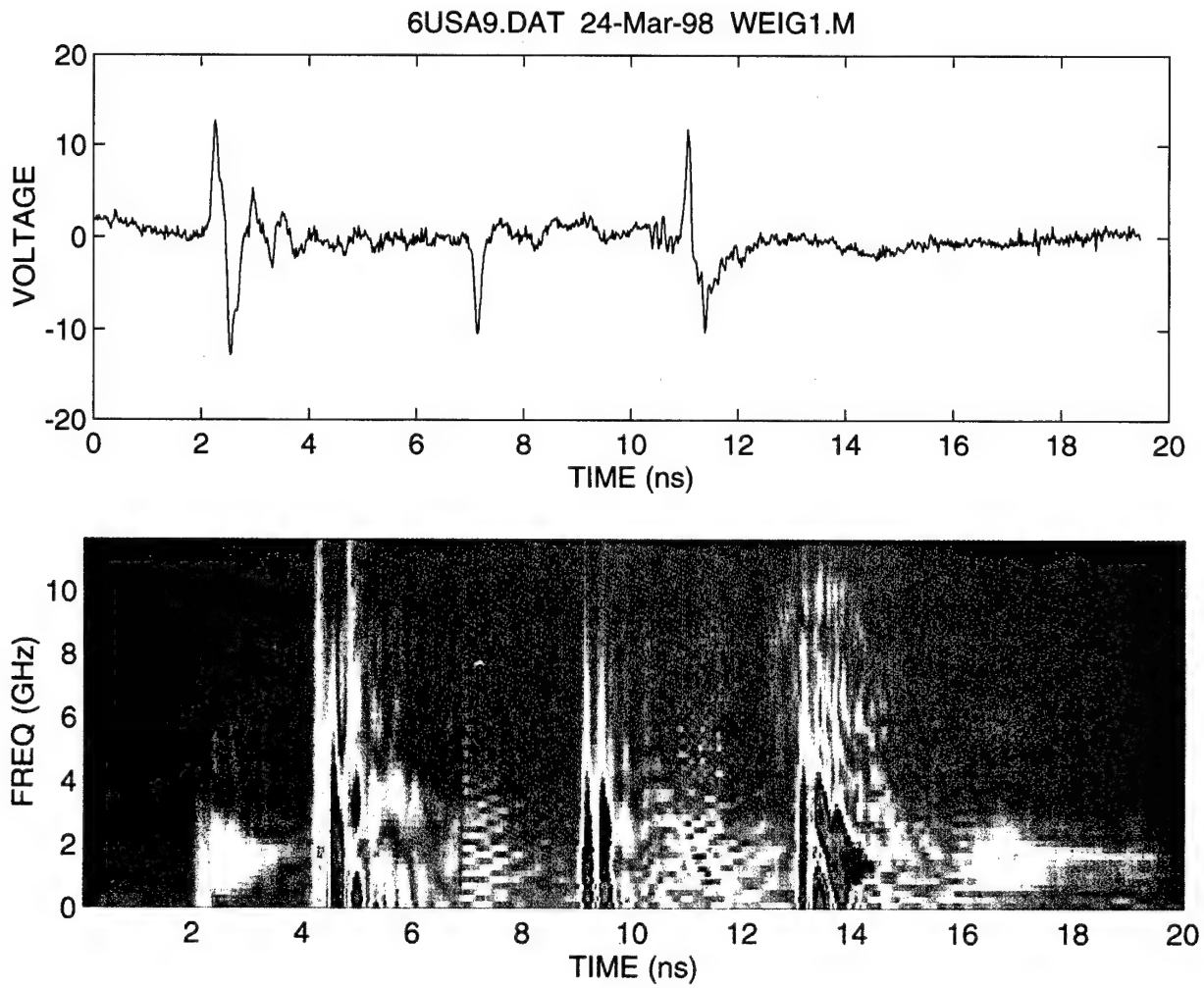
Figure 12. Raw data and Wigner algorithm for 6USA9(45°).
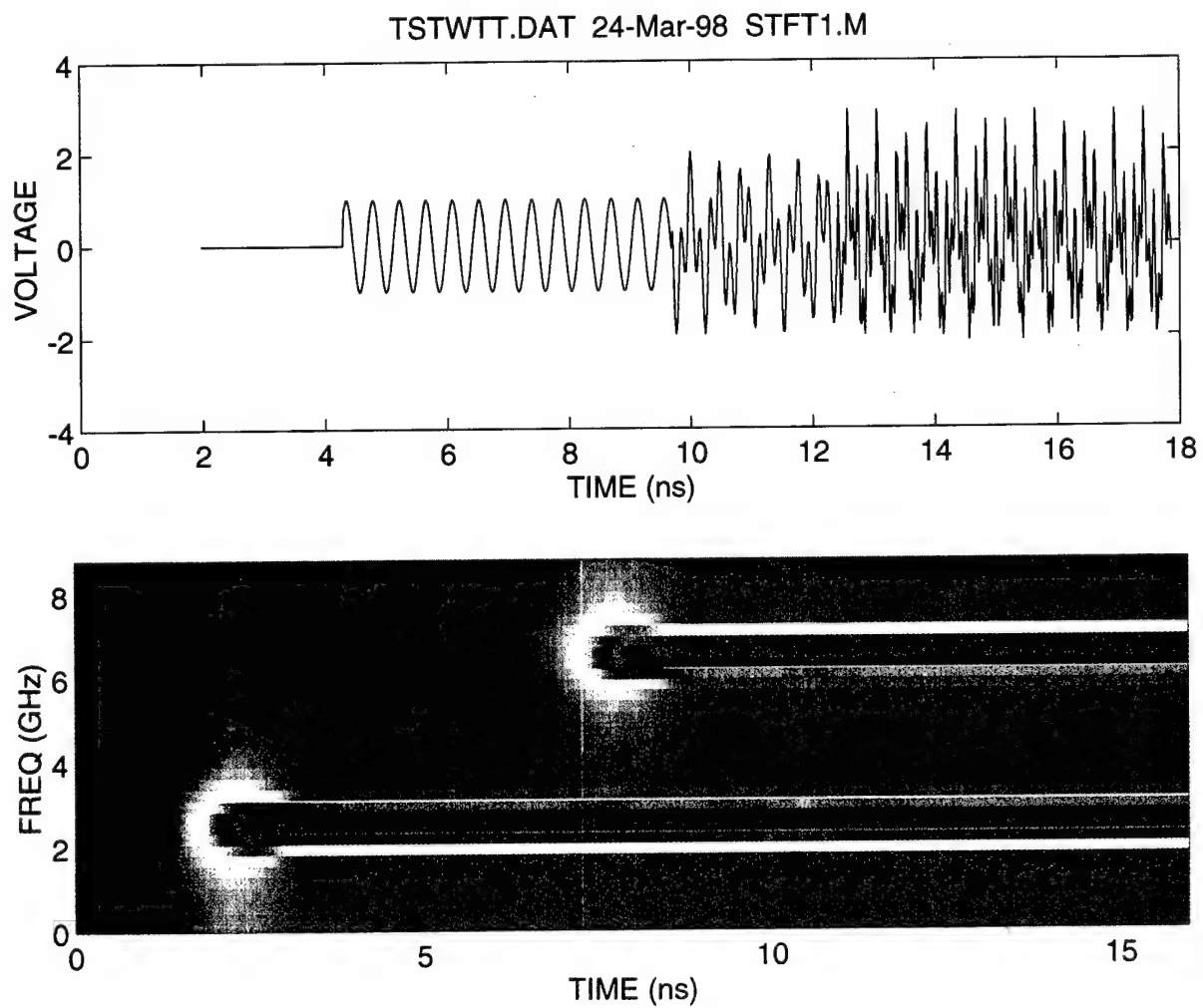
Figure 13. Raw waveform and STFT image of theoretical data. (impulse at 1.1 ns; 2.3 GHz starting 4.3 ns; 6.7 GHz starting 9.7 ns; 12.4 GHz starting at 12.4 ns).
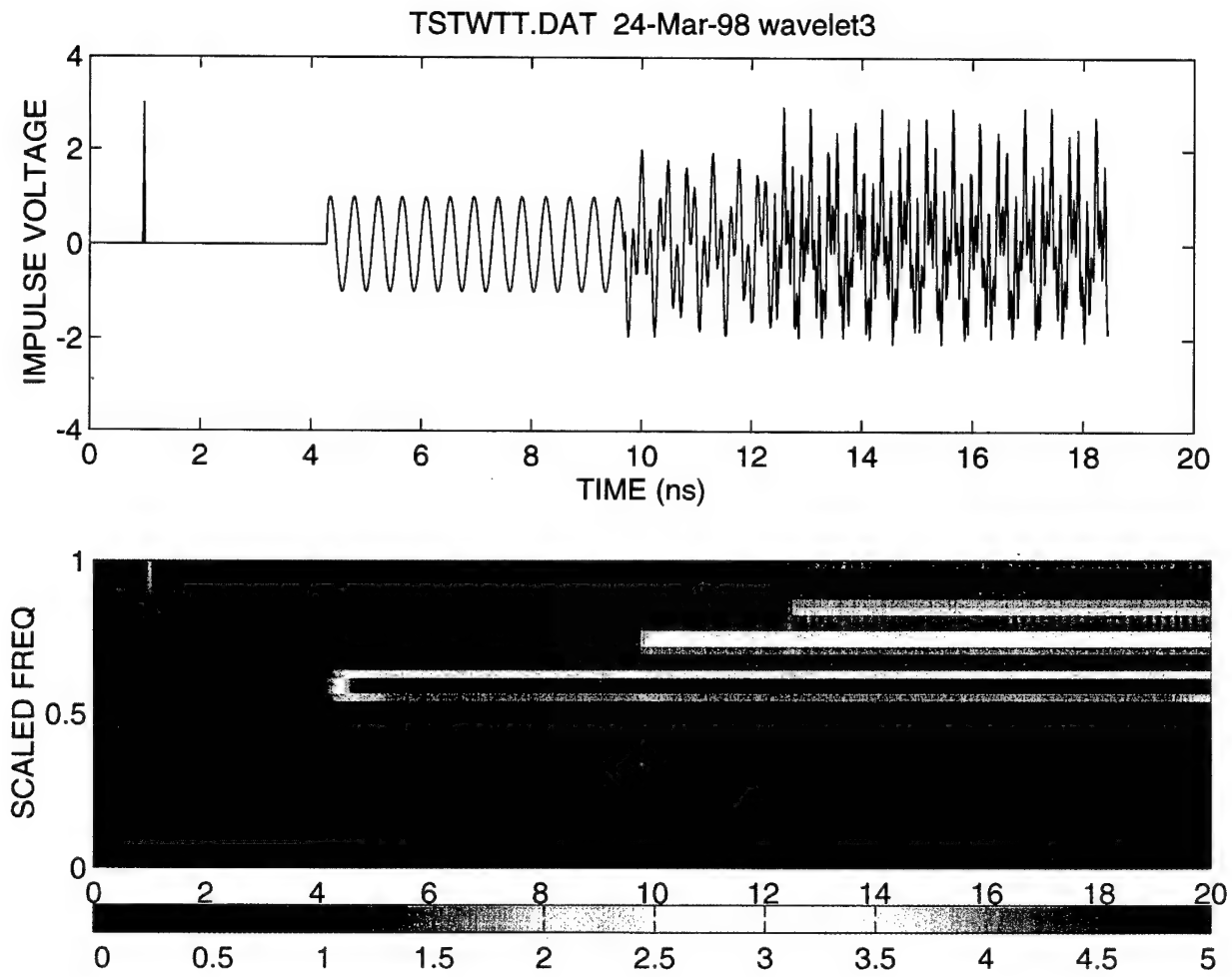
Figure 14. Raw waveform and wavelet image of theoretical data. (impulse at 1.1 ns; 2.3 GHz starting 4.3 ns; 6.7 GHz starting 9.7 ns; 12.4 GHz starting at 12.4 ns).
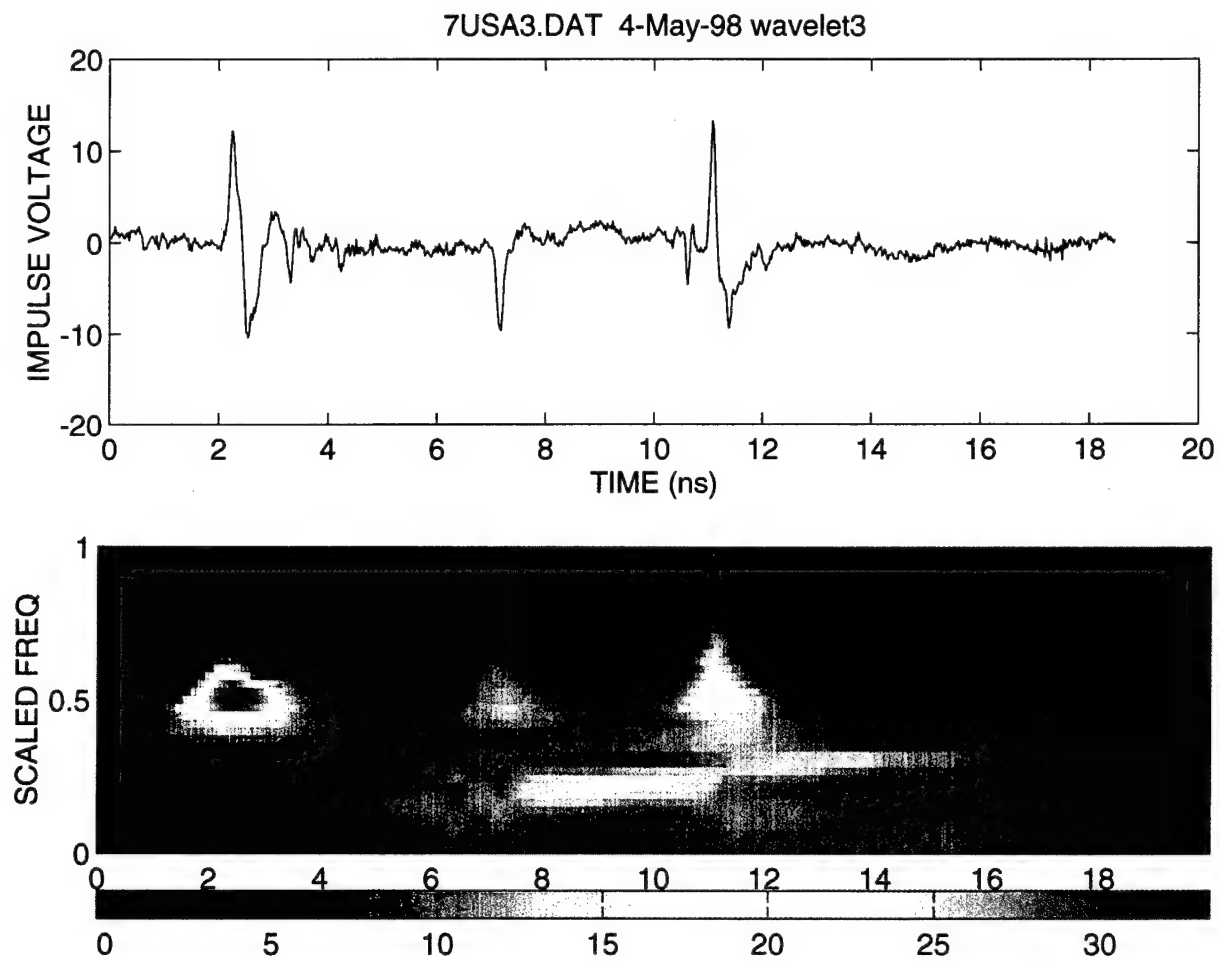
Figure 15. Raw data and wavelet image of 7USA3 (45°).
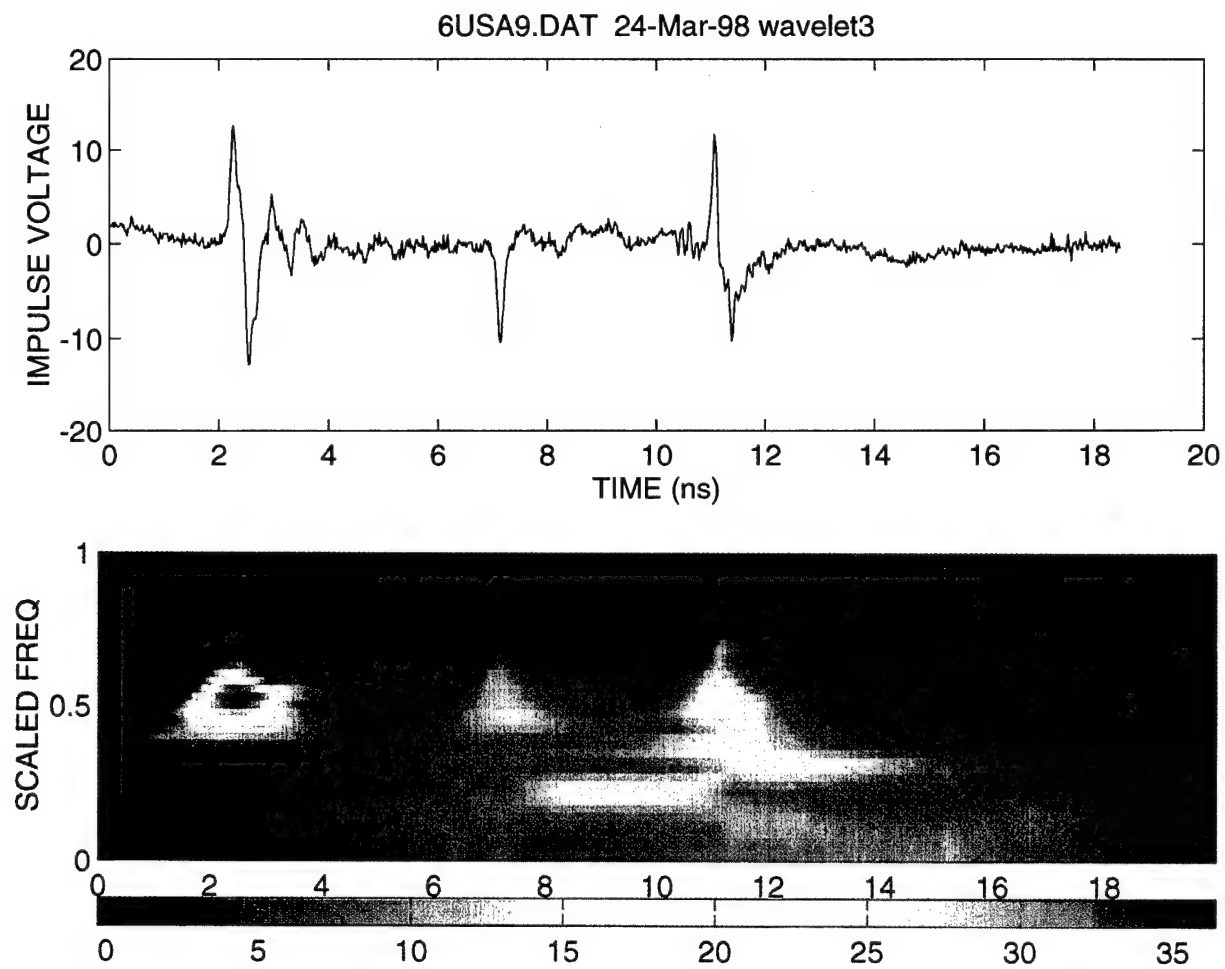
Figure 16. Raw data and wavelet image of 6USA9 (45°).
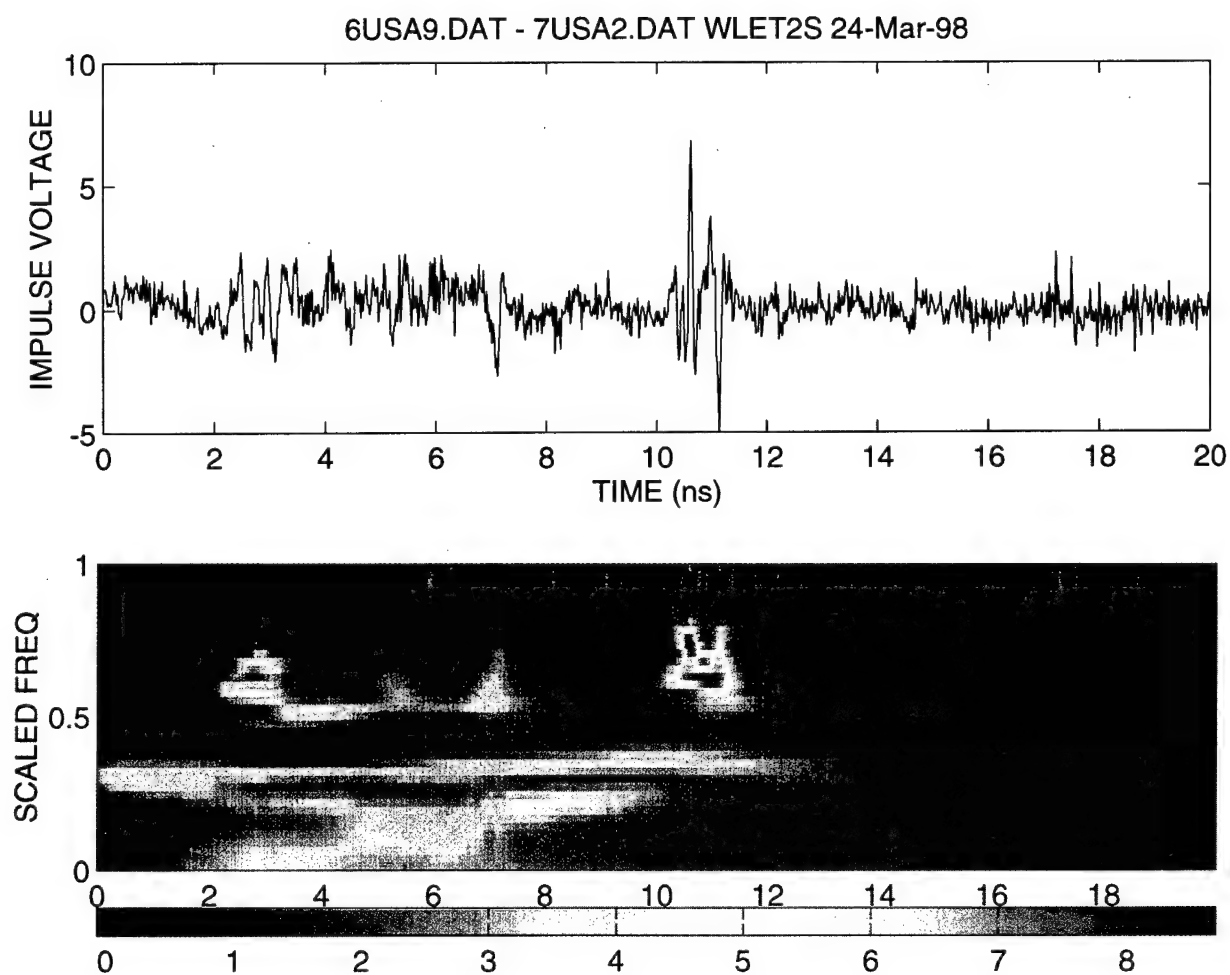
Figure 17. Raw data and wavelet image of 6USA9-7USA2 (subtraction).
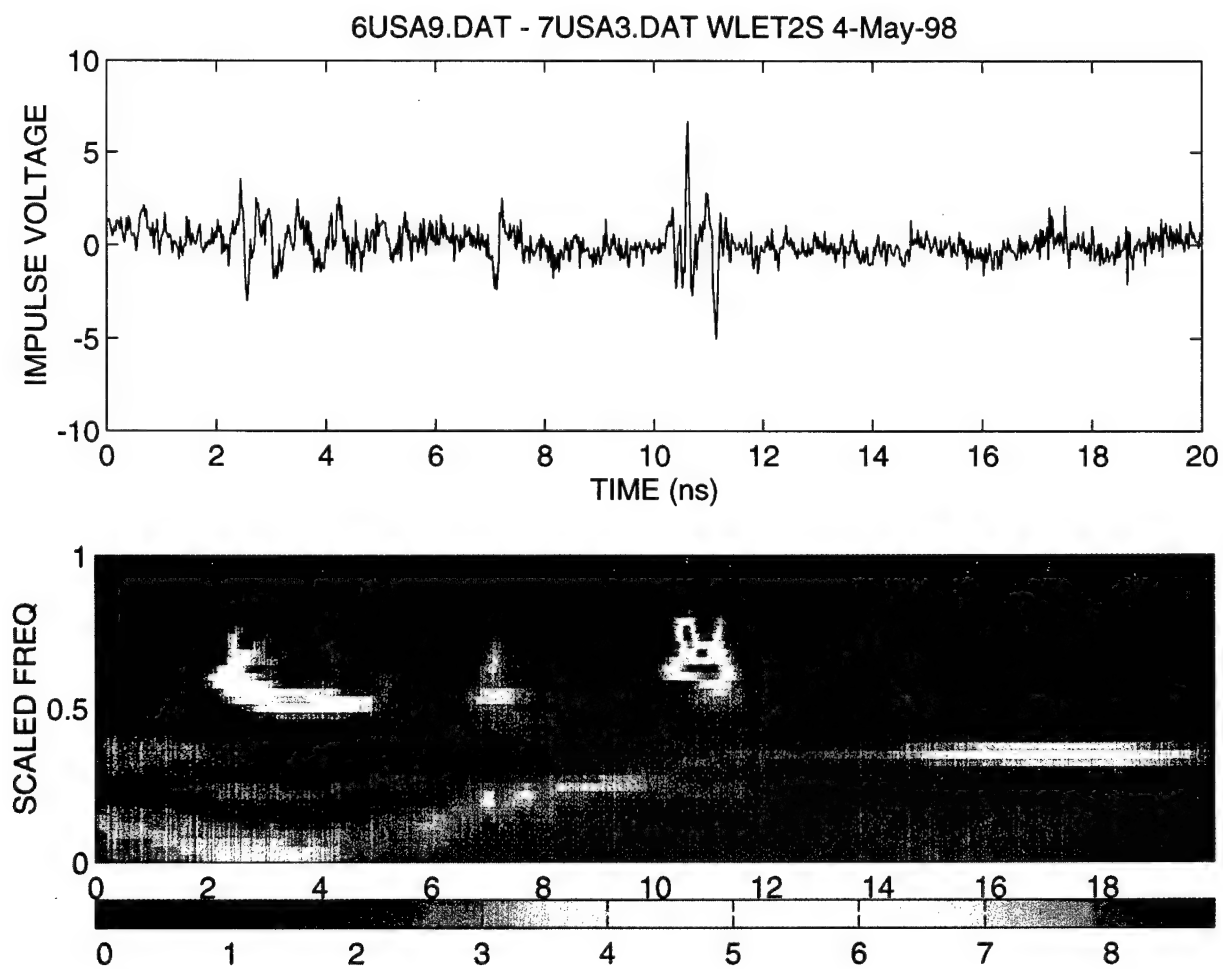
Figure 18. Raw data and wavelet image of 6USA9-7USA3 (subtraction).
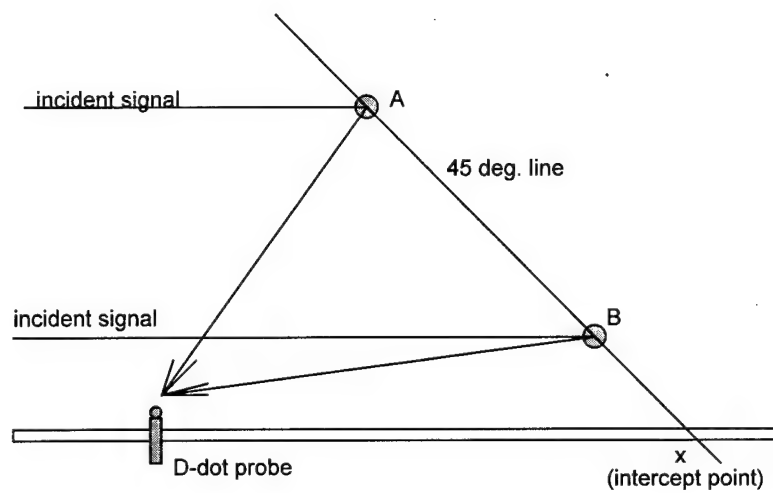
Figure 19. Geometry for 45° bistatic interaction.
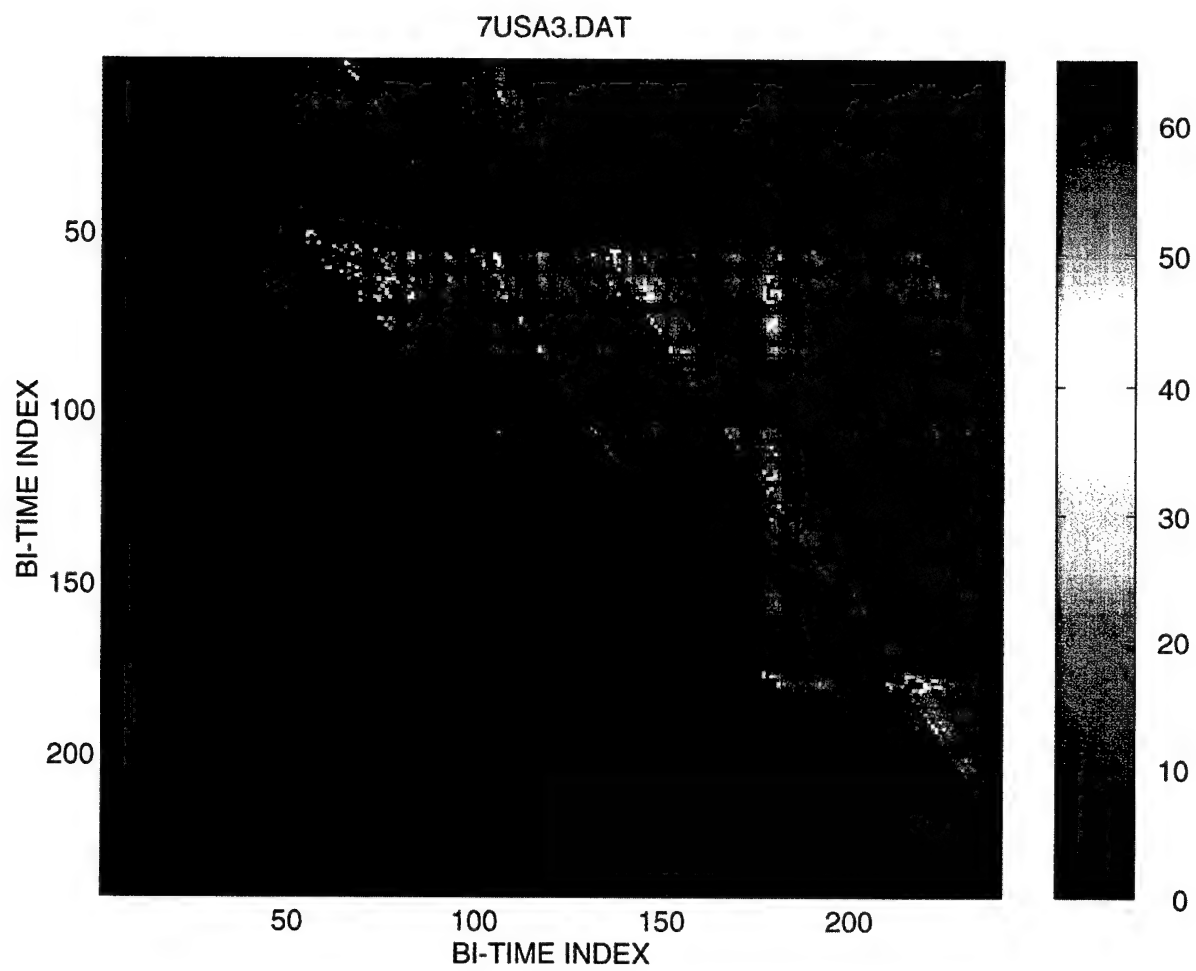
Figure 20. Bispectral image of 7USA3.
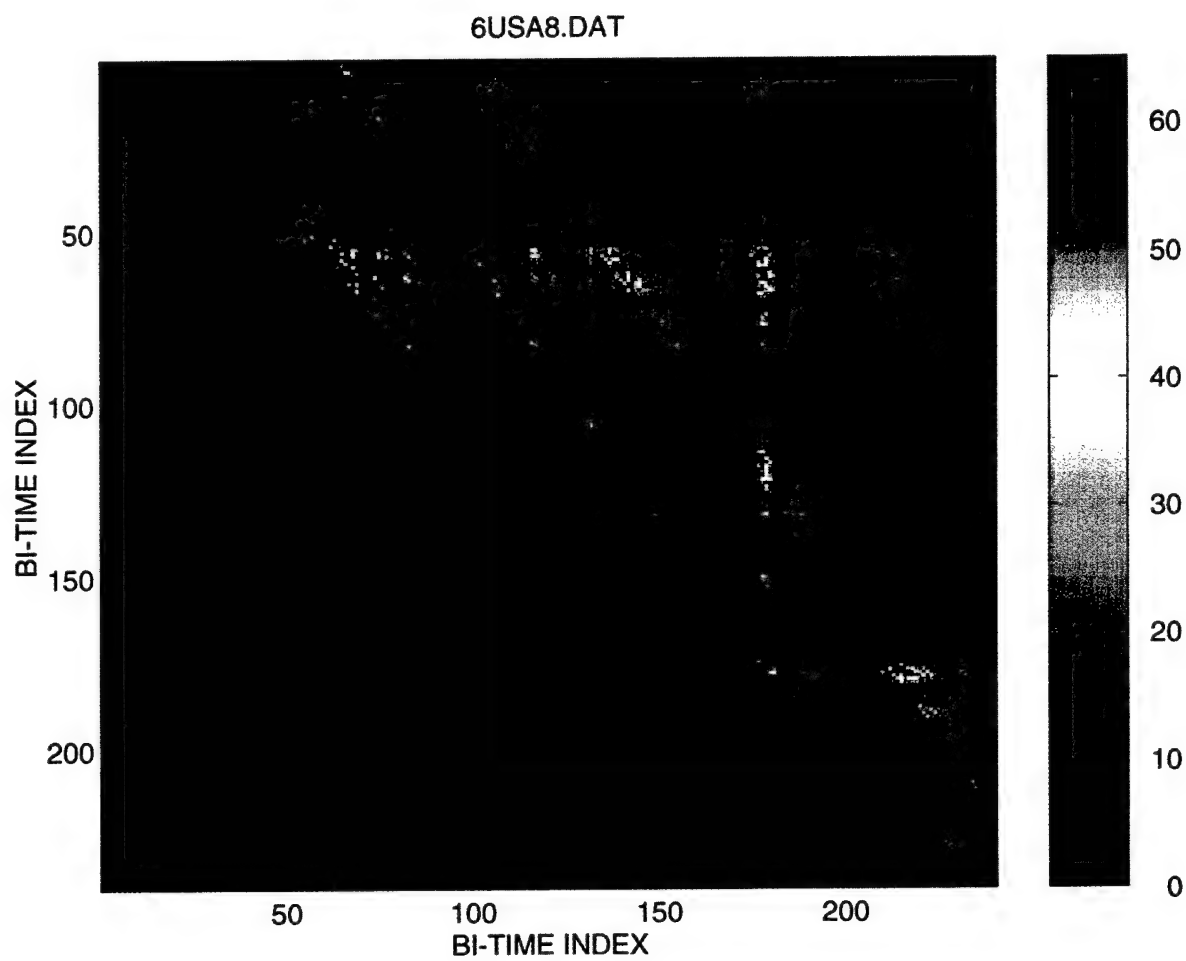
Figure 21. Bispectral image of 6USA8.

7USA2.DAT
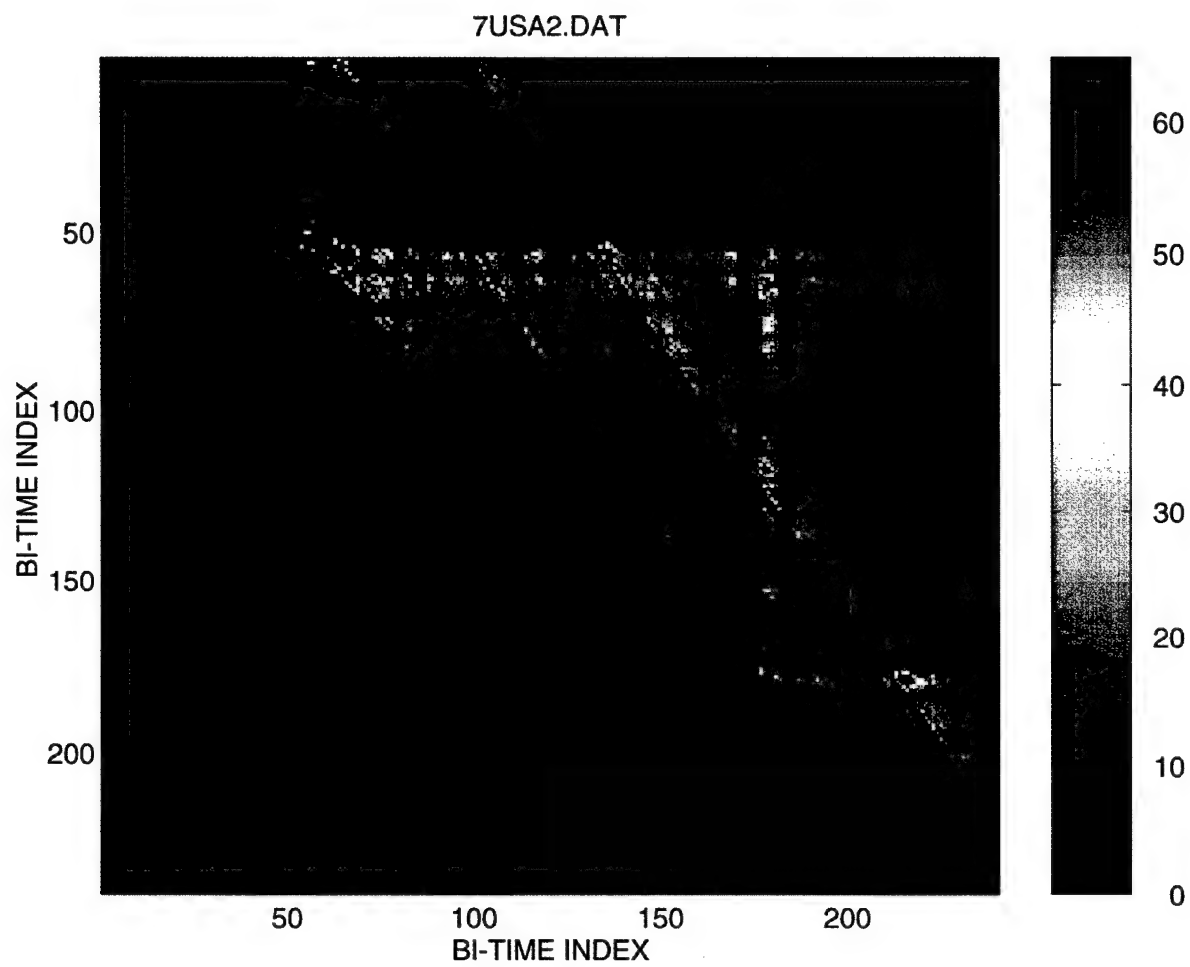
Figure 22. Bispectral image of 7USA2.

6USA9.DAT
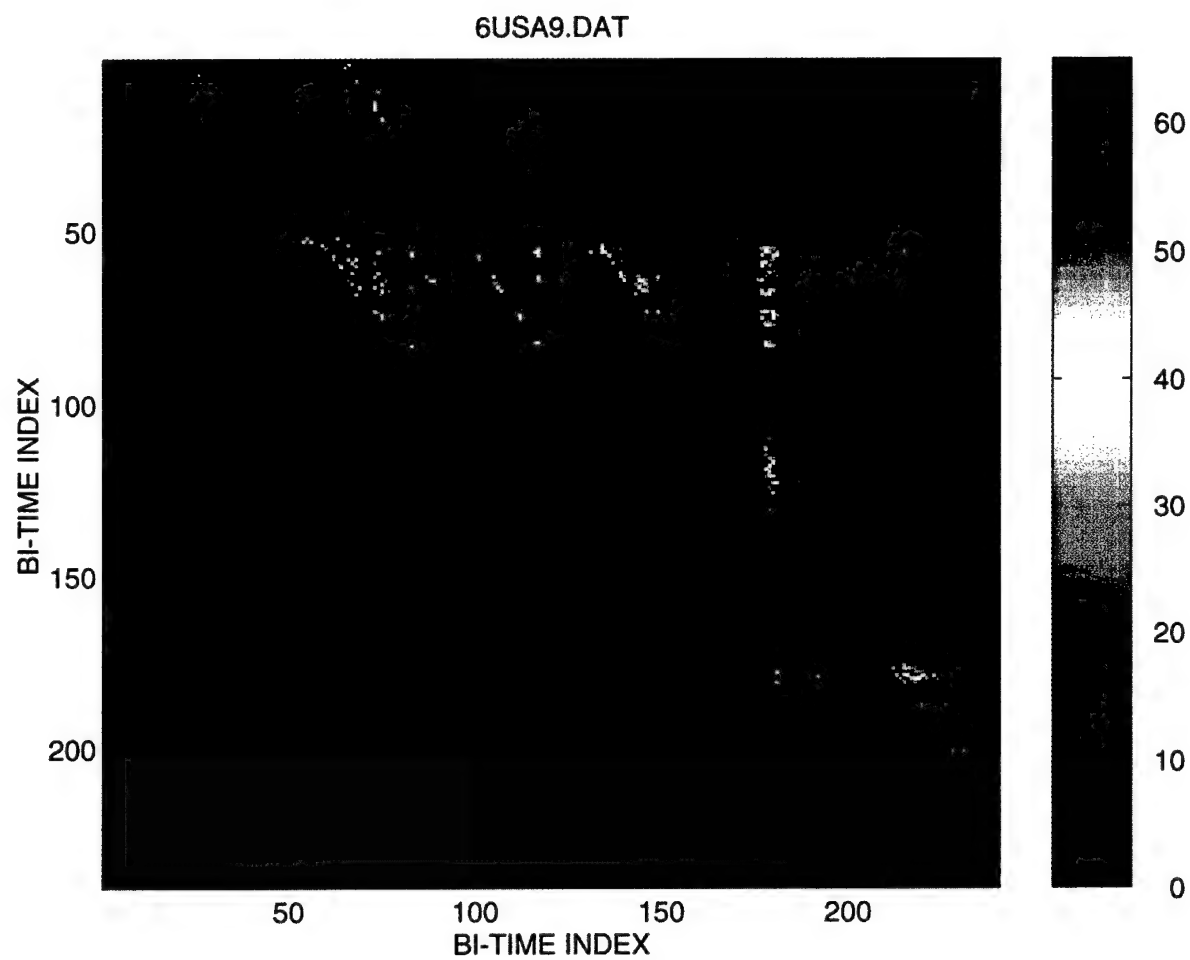
Figure 23. Bispectral image of 6USA9.

Electromagnetic Inversion

A. Peter M. Zwamborn
TNO Physics and Electronics Laboratory
Oude Waalsdorperweg 63
P.O. Box 96864
2509 JG 's Gravenhage
The Netherlands

1. INTRODUCTION

1.1 The Land Mine Pollution Problem

The pollution of areas with large quantities of anti-tank and anti-personnel land mines, especially in countries of former armed conflicts, like Afghanistan, Angola, Cambodia, Iraq, Kuwait, Somalia, Vietnam and Yugoslavia, is a major problem. According to United Nations estimates, the number of uncharted buried anti-personnel mines exceeds 100 million, in over 60 countries around the world. The rate of new mines being laid is about one million per year, which surpasses the number of mines cleared by a factor of twenty. Some 2 million mines have been deposited in the war-torn areas of former Yugoslavia alone. Whole areas of countries, especially Cambodia, have been severely held back from further development [1].

According to the Mine Clearance Planning Agency in Afghanistan, over a period of 15 years an estimated 20,000 civilians have been killed and 400,000 wounded by land mines in that country. The current rate is 4000 killed and another 4000 wounded annually, world-wide. Injuries are horrific and usually result in amputation; returning refugees to conflict zones often find minefields in previously farmed lands, and usually have to clear the land themselves. Methods used at present for locating and clearing mines are painstaking, costly, time consuming and highly dangerous. Those methods include the use of sniffer dogs, magnetic mine detection aids (e.g. metal detectors) and manual probing. These methods are very slow and involve teams of two working their way along rows as narrow as 1 m across. For instance, a team of 30 men with dogs is able to cover only 2000 squared meters per day. The costs of such clearance is reported to lie between $ 200 (US dollars) and $ 1000 per mine. After the food problem, the so called land mine pollution problem is seen as the biggest humanitarian problem in the world [1].

Initial moves have been made towards a world-wide ban of land mines. In December 1993, the United Nations General Assembly passed a non-binding resolution calling for such a ban. In the 1981 United Nations Convention, some rules governing the use of land mines (considering for example the automatic neutralization of land mines and the obligation to record pre-planned land minefields by means of maps) have been agreed upon. This international law regulating the use of land mines, the 1981 Land Mines Protocol, at present only regulates the use of land mines in wars, but not in internal conflicts, and has been ratified by only 39 countries. Meanwhile, new minefields have been created in Georgia, Armenia and Tajikistan, for example, as well as in the territories of former Yugoslavia.

In May 1996, a United Nations Review Conference of the 1981 Convention finished in Geneva. One of the main topics was land mines. Some additional rules to those of 1981 were agreed upon. Internal conflicts are now also covered by the convention. It was further agreed that land mines must have a self-deactivation device and that they must contain at least 8 g of iron, to be detectable by the current types of mine detectors.

The above discussion on the land mine pollution problem demonstrates the need to develop new technologies to increase the efficiency and to reduce the costs of mine clearing operations.

1.2 Electromagnetic Inversion

The detection of objects buried in the ground is in general difficult. An even bigger problem is discrimination or classification of the buried object. The first problem is the ground itself. It is usually very inhomogeneous and has a complicated layered structure, often containing rocks and voids. Moreover, many other objects like metal cans can be present in the ground. Without a reliable identification method, the false-alarm rate of a ground-penetrating-radar system would be so high that the cost of clearing of a minefield would be prohibitive.

The objective of this lecture on inverse methods is to present some methods to carry out detection of buried objects. Subsequently, the reconstructed data is available to dedicated pattern recognition algorithms to obtain classification of the detected object.

In general the electromagnetic data is obtained by using fixed-frequency systems operating at a single or multiple frequencies or ultra-wideband systems. The term ultra-wideband is used in all situations where one deals with pulses of extremely short duration. A pulse of almost zero duration, which approximates a delta function, contains almost all frequencies. Hence, short pulses are ultra-wideband pulses. It has been claimed by some authors that UWB systems have many advantages compared to fixed-frequency systems when used for probing the ground [2], [3]. In a UWB system, like a ground-penetrating-radar system, a large amount of individual frequencies are applied towards the object of interest.

1.3 The EM-Inversion Methods Presented in this Lecture

The first method that is presented is denoted as "Microwave Image Reconstruction Methods" by S. Primak *et. al.* This method uses a special mapping of gathered data and in this paper a tutorial type overview of microwave tomographic imaging is given. The second method that is presented is denoted as "Two-Dimensional Inverse Profiling: Nonlinear Optimization and Embedding" by A. Tijhuis *et. al.* In this

paper, a method is presented to solve nonlinear inverse scattering problems. This approach is based on the availability of efficient iterative solvers to carry out the electromagnetic computations. The last method discussed in this lecture is denoted as "Non-linear Inversion Based on Contrast Source Gradients" by van den Berg *et. al.* This method is an algorithm for reconstructing the complex index of refraction of a bounded object. Also, this method incooperates efficient iterative solvers.

REFERENCES

1. EUREL,
   International Conference on the detection of abandoned land mines: a humanitarian imperative seeking a technical solution,
   Edinburgh, October 7th-9th 1996,IEE conference publication 431,
   Institute of Electrical Engineers,London,ISBN 0 5296 669 5.

2. Baum, C.E.,
   Signature-based target identification and pattern recognition,
   IEEE Antennas and Propagation Magazine,
   vol. 36, no. 3, pp. 44-51, 1994.

3. Baum, C.E. and E.J. Rothwell, K.M. Chen and D.P.Nyquist,
   The Singularity Expansion Method and its application to target identification,
   Proc. of the IEEE, vol. 79, no. 10, pp. 1481-1492, 1991.

Microwave Image Reconstruction Methods

Sergey Primak, Joe LoVetri, and Beibei Zhang
Department of Electrical and Computer Engineering
The University of Western Ontario
London, Ontario, Canada N6A 5B9

1. SUMMARY

In this paper we give a tutorial type overview of microwave tomographic imaging, or inverse synthetic aperture ISAR (and SAR) imaging, of perfectly conducting bodies by considering the simplest high-frequency model of scattering and we review methods allowing one to obtain these images.

2. IDEAL POINT SCATTERER MODEL AND ISAR

ISAR (SAR) imaging provides a way to map data gathered at multiple frequencies and aspect angles into a two-dimensional image, where the points correspond to $x - y$ coordinates in space (called *down-range* and *cross-range* respectively) [3]. An intuitive explanation based on ideal point scatterers will be given in this section.

Consider a monochromatic electromagnetic plane wave, having a frequency ω, incident at an angle φ_i, on a target containing an ideal scattering centre of strength A at position (x_0, y_0) (see Fig. 1). The incident plane wave is represented as

$$E_{inc} = \exp(-j\mathbf{k} \cdot \mathbf{r} + j\omega t) \qquad (1)$$

where $\mathbf{k} = k(\hat{a}_x \cos\varphi_i + \hat{a}_y \sin\varphi_i) = \hat{a}_x k_x + \hat{a}_y k_y$ is the wave vector with $k = \omega/c$ (c is the speed of propagation), $\mathbf{r} = \hat{a}_x x + \hat{a}_y y$ is the position vector in the plane, $\hat{a}_x$ and $\hat{a}_y$ are the unit vectors in x and y directions, respectively. The resulting backscattered field can be written as [3]

$$E_{sc} = A\exp(-2jk_x x_0 - 2jk_y y_0)\exp(j\mathbf{k} \cdot \mathbf{r} + j\omega t) \qquad (2)$$

Dividing the scattered field by the incident field and multiplying by $\exp(-2\mathbf{k} \cdot \mathbf{r})$ one obtains

$$S(\omega, \varphi_i) = A\exp(-2jkx_0\cos\varphi_i - 2jky_0\sin\varphi_i) \qquad (3)$$

If we take the 2-D inverse Fourier transform of (2) with respect to $2k_x$ and $2k_y$ we obtain

$$s(x, y) = A\delta(x - x_0)\delta(y - y_0) \qquad (4)$$
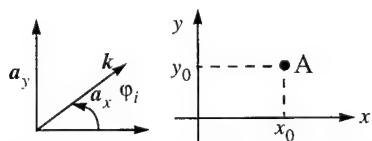
Figure 1. Incident wave vector and ideal scattering centre.

This is the essential idea behind ISAR processing: ideal scattering centres correspond to ideal points in the ISAR image [3, 4].

3. METHODS OF IMAGE CONSTRUCTION

As we've seen, but not proved, the scatterer geometry is simply the 2-D inverse Fourier Transform of the reflectivity of the object in the $k_x - k_y$ plane. Methods of image construction from measurements, based on the direct application of the 2-D inverse Fourier Transform as well as alternative methods, making use of the so called *Central Slice Theorem*, will now be discussed.

3.1. 2-D inverse Fourier Transform

As has just been discussed, an image of the geometry of a scatterer can be constructed by an inverse 2-D Fourier Transform of the scattered signal in the frequency domain, that is [5, 6]

$$s(x, y) = \frac{1}{\pi^2}\int_{R^2} S(k_x, k_y)\exp[2j(k_x x + k_y y)]dk_x dk_y \qquad (5)$$

This expression gives us a direct method of recovering the image from the measurements. Unfortunately, the discrete version of (5) requires the uniform rectangular sampling of information in the $k_x - k_y$ domain, while the measurements are usually taken in the $\omega - \varphi_i$ domain, which is non-uniform in the $k_x - k_y$ domain. Thus some kind of interpolation has to be used (see [5] for details).

3.2. Filtering and backprojection

An alternative method will now be described (see [6, 7, 11]). Using the fact that $k_x = (\omega/c)\cos\varphi_i$, and $k_y = (\omega/c)\sin\varphi_i$ one can rewrite integral (5) as

$$s(x, y) = \frac{1}{\pi^2}\int_{R^2} S(k_x, k_y)\exp[2j(k_x x + k_y y)]dk_x dk_y$$
$$= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{0}^{\pi} S_{\varphi_i}(\omega)e^{2j\omega(x\cos\varphi_i + y\sin\varphi_i)}|\omega|d\omega d\varphi_i \qquad (6)$$

where $S_{\varphi_i}(\omega) = S(\omega, \varphi_i)$ is the slice of the frequency domain image (frequency response of the of the target) taken at the angle φ_i. Using $\rho = \sqrt{x^2 + y^2}$ and $\beta = \operatorname{atan}(y/x)$, we get

$$s(\rho, \beta) = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{0}^{\pi} S_{\varphi_i}(\omega)e^{2j\omega\rho(\cos\beta\cos\varphi_i + \sin\beta\sin\varphi_i)}|\omega|d\omega d\varphi_i$$
$$= \frac{1}{\pi^2}\int_{0}^{\pi}d\varphi_i \int_{-\infty}^{\infty} S_{\varphi_i}(\omega)e^{2j\omega\rho\cos(\beta - \varphi_i)}|\omega|d\omega \qquad (7)$$

Therefore the shape of the scatterer can be obtained by first filtering the frequency domain slices, that is obtain the filtered signal

$$\hat{s}_{\varphi_i}(\rho\cos(\beta - \varphi_i)) = \frac{1}{2\pi}\int_{-\infty}^{\infty}|\omega|S_{\varphi_i}(\omega)e^{2j\omega\rho\cos(\beta - \varphi_i)}d\omega \tag{8}$$

$$= F_{2\rho\cos(\beta - \varphi_i)}^{-1}[|\omega|S_{\varphi_i}(\omega)]$$

This is just the 1-D inverse Fourier transform (using the spatial parameter $2\rho\cos(\beta - \varphi_i)$) of the frequency domain slice, taken at angle φ_i and multiplied by $|\omega|$. Alternatively, it is the spatial domain slice convolved with $h(t) = F^{-1}[|\omega|]$ [6].

The next step is called *back-projection* of the filtered slices:

$$s(\rho, \beta) = \frac{2}{\pi}\int_0^{\pi} d\varphi_i \hat{s}_{\varphi_i}(\rho\cos(\beta - \varphi_i)) \tag{9}$$

This states that the reconstructed function $s(\rho, \beta)$ is the result of averaging the signal $\hat{s}_{\varphi_i}(\rho\cos(\beta - \varphi_i))$ with respect to φ_i, which, in turn, is the back-projection of the signal $\hat{s}_{\varphi_i}(x)$ along the line in the same direction in which the projection function is obtained (see Fig. 2). Thus, the reconstructed pixel is the averaged back-projection of the measurements, taken all around the object. Since the filtered version of the measurements is used, this algorithm is called *filtered back-projection* [6]. From equation (9), the discrete approximation of the reconstructed function $s(\rho, \beta)$ can be obtained as follows

$$s(\rho, \beta) = \int_0^{\pi} d\varphi_i \hat{s}_{\varphi_i}(\rho\cos(\beta - \varphi_i))$$
$$= \frac{\pi}{N}\sum_{n=0}^{N-1}\hat{s}_n(\rho\cos(\beta - \pi n/N)) \tag{10}$$

which can be performed sequentially as a new measurement is obtained. Thus, the filtered back-projection algorithm employs only a series of 1-D inverse Fourier transforms and does not require the complete data set to start reconstruction. This makes the algorithm the best choice for reconstruction of the function from its projections [6, 13].
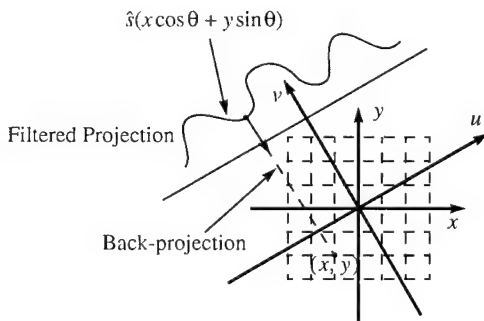
Figure 2. Back-projection algorithm

3.3. *Back-projection and filtering*

The filtered back-projection is not the only way to reconstruct the desired function from the measurements. In fact, equation (5) can be rewritten as

$$s(x, y) = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_0^{\pi}S_{\varphi_i}(\omega)e^{2j\omega(x\cos\varphi_i + y\sin\varphi_i)}|\omega|\,d\omega d\varphi_i \tag{11}$$

or, using the variables β and ρ, as

$$s(\beta, \rho) = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_0^{\pi}S_{\varphi_i}(\omega)e^{2j\omega\rho(\cos\beta\cos\varphi_i + \sin\beta\sin\varphi_i)}|\omega|\,d\omega d\varphi_i$$

$$= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_0^{\pi}S_{\varphi_i}(\omega)e^{2j\omega\rho\cos(\beta - \varphi_i)}|\omega|\,d\omega d\varphi_i \tag{12}$$

$$= \frac{2}{\pi}F^{-2}(|\omega|)*\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_0^{\pi}S_{\varphi_i}(\omega)e^{2j\omega\rho\cos(\beta - \varphi_i)}d\varphi_i$$

and, finally,

$$s(\beta, \rho) = \frac{2}{\pi}F^{-2}(|\omega|)*\int_0^{\pi}s_{\varphi_i}(\rho\cos(\beta - \varphi_i))d\varphi_i \tag{13}$$

where $*$ denotes 2-D convolution. The final integral term in (12) represents the back-projection of the signal, restored from the frequency domain measurements without any filtering, *i.e.*

$$s_{\varphi_i}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}S_{\varphi_i}(\omega)e^{2j\omega x}d\omega \tag{14}$$

but to restore the true function $s(\rho, \varphi_i)$ one has to convolve the result of the back-projection with a *point-spread function* [6]

$$F^{-2}(|\omega|) = \rho^{-1}, \quad \rho = \sqrt{x^2 + y^2} \tag{15}$$

As we can see, this algorithms allows the reconstruction of the image in two steps, one of which requires a two-dimensional convolution with a singular function. Where only a rough image of the body is required, the convolution may be omitted.

4. PHYSICS-BASED SCATTERING CENTRE MODEL

The previous considerations were based on the following assumptions:

1) that scattering from a complex object can be thought of as scattering from a relatively small number of ideal point scatterers [3, 4]; and that

2) measurements are available for all incident angles in the $x-y$ plane or, at least, for a discrete set of angles, covering the range $[0, 2\pi]$, [6, 7].

In practice, any of these assumptions may become invalid. Due to the limited frequency content of the excitation, especially taking into account its base-band nature, the real physical object cannot be represented as a set of ideal point scatterers [7, 8]. In fact, the reflection coefficients depend on frequency and angle of incident. Also, if a wide range of angles is used some scattering centres may become invisible for some of the aspect angles. Cavities, ducts, and other structures present in real targets produce scattered signals that have a very strong frequency dependence which cannot be "focused" on the ISAR (SAR) image.

For realistic targets it is almost impossible to obtain the information about the scattered waves for a wide range of aspect angles. This is why many authors have considered microwave imaging using the back-projection algorithm as a narrow-band (with respect to the angular variable) technique.

Another physical phenomena, inherently presented, is the dependence of the scattering parameters on the polarization of the incident wave.

In this section we will present more realistic models of the targets, based on the geometrical theory of diffraction (GTD) [9].

4.1. GTD scattering model

The GTD scattering model is valid under the assumption that the wavelength of the incident excitation is small relative to the targets extent [9]. It represents the scattered field as originating from a set of M discrete scattering centres located at points (x_m, y_m), $1 \leq m \leq M$, i.e.

$$E^s(\omega, \varphi_i, t) \approx \frac{E_0}{c_0 t} e^{j\omega t} \sum_{m=1}^{M} S_m(\omega, \varphi_i) \exp[-j\omega T_m] \quad (16)$$

where ω is the angular frequency, c_0 is the propagation velocity,

$$T_m = \frac{2}{c_0}(x_m \cos\varphi_i + y_m \sin\varphi_i) \quad (17)$$

is the round trip delay of the m-th scattering centre, and M is the number of scattering centres. The amplitude $S_m(\omega, \varphi_i)$ is a frequency and angle dependent coefficient determined by the geometry, composition, and orientation of the scattering mechanism. The normalized field for a given polarization (suppressing the $e^{j\omega t}$ dependence) is then

$$\rho(\omega, \varphi_i) = \sum_{m=1}^{M} S_m(\omega, \varphi_i) e^{-2jk(x_m \cos\varphi_i + y_m \sin\varphi_i)} \quad (18)$$

For frequency dependence, GTD predicts that the scattering amplitude is of the form $(j\omega)^{\alpha_m}$ where α_m is an integer multiple of 0.5. Table 1 shows the relation between different values of the parameter α_m and the geometry of the scatterer [10].

Table 1.   Relation between α and scatterer geometry

Value of α	Canonical scatterers
1	Flat plate at broadside; dihedral
1/2	Singly curved surface reflection
0	point scatterer; doubly curved surface reflection straight edge specular
-1/2	curved edge diffraction
-1	corner diffraction

The angle dependence of the magnitude of the scattered field is a much more complicated question. Most man-made targets, containing sharp edges can be characterized by a sinc-like aspect dependence of the scattering coefficient. In practice, especially, for the narrow-band angle ISAR (SAR), this dependence can be accurately approximated by the exponential, i.e. [10]

$$S(\omega, \varphi_i) = A_m(j\omega)^{\alpha_m} e^{\beta\varphi_i} \quad (19)$$

Thus, GTD predicts the following model for the scattering from large targets

$$\rho(\omega, \varphi_i) = \sum_{m=1}^{M} A_m(j\omega)^{\alpha_m} e^{\beta\varphi_i} e^{-2j\frac{\omega}{c_0}(x_m \cos\varphi_i + y_m \sin\varphi_i)} \quad (20)$$

which sometimes results in ISAR (SAR) image of a target different from those obtained using PO approximation.

4.2. ISAR image of the point scatterer from narrow band of angles

If the measurements are performed in a narrow range of incident angles, say $-\alpha_{max} < \varphi < \alpha_{max}$, and in the frequency range $-\omega_{max} < \omega < \omega_{max}$, then

$$|k_x| \leq \frac{\omega_{max}}{c_0}, \quad |k_y| \leq \frac{\omega_{max}}{c_0}\sin\alpha \quad (21)$$
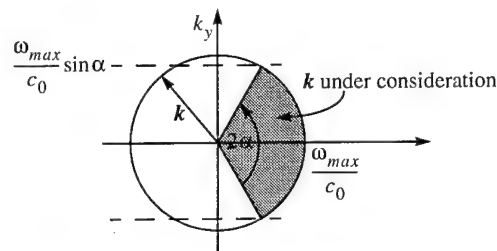
Figure 3. A narrow-band angular $k_x - k_y$ plane

and the image of the scatterer is given by

$$s(x, y) = \frac{1}{\pi^2} \int_{-\frac{\omega_{max}}{c_0}}^{\frac{\omega_{max}}{c_0}} \int_{-\frac{\omega_{max}}{c_0}\sin\alpha}^{\frac{\omega_{max}}{c_0}\sin\alpha} e^{2j[k_x(x-x_0)+k_y(y-y_0)]} dk_x dk_y$$

$$= \frac{1}{\pi^2} \frac{\sin\left[\frac{\omega_{max}}{c_0}(x-x_0)\right]}{\frac{\omega_{max}}{c_0}(x-x_0)} \frac{\sin\left[\frac{\omega_{max}}{c_0}(y-y_0)\sin\alpha\right]}{\frac{\omega_{max}}{c_0}(y-y_0)\sin\alpha} \quad (22)$$

which approximately represents an ellipse in the ISAR plane.

4.3. Resonant scattering

As was mentioned before, the low frequencies in the incident waveform can excite the resonance of the target [14]. This effect can be described by an extra term in the frequency response of the target. If we assume that the resonant response is localized at the point (x_0, y_0) and the frequency of oscillations is ω_0, with a damping factor is σ, and an angle dependence which is proportional to a decaying exponential

term $\exp(\beta\varphi_i)$, then the extra term due to the resonance can be written as

$$S_r(\omega, \varphi_i) = \exp(\beta\varphi_i)\frac{\exp\left(-2j\frac{\omega}{c}x_0\cos\varphi_i - 2j\frac{\omega}{c}y_0\sin\varphi_i\right)}{\sigma + j(\omega - \omega_0)} \quad (23)$$

If we assume also that data are gathered over a narrow range of angles $-\alpha < \varphi_i < \alpha$, equation (23) can be rewritten as

$$S_r(\omega, \varphi_i) \approx \frac{\exp\left(-2j\frac{\omega}{c}x_0 - 2j\frac{\omega}{c}y_0\varphi_i\right)}{\sigma + j(\omega - \omega_0)} \quad (24)$$

thus, the image in the ISAR plan will be given

$$\int_{-\omega_{max}}^{\omega_{max}}\int_{-\alpha}^{\alpha} |\omega|\frac{\exp\left(-2j\frac{\omega}{c}x_0 - 2j\frac{\omega}{c}\varphi_i\right)}{\sigma + j(\omega - \omega_0)}\exp[2j(k\cdot r)]d\omega d\varphi_i$$

$$\approx \int_{-\omega_{max}}^{\omega_{max}}\int_{-\alpha}^{\alpha} |\omega|\frac{\exp\left\{2j\frac{\omega}{c_0}[(x-x_0) + (y-y_0)\varphi_i]\right\}}{\sigma + j(\omega - \omega_0)}d\omega d\varphi_i \quad (25)$$

$$= C\frac{\sin\left[\frac{\omega}{c_0}(y-y_0)\sin\alpha\right]}{\frac{\omega}{c_0}(y-y_0)\sin\alpha} \cdot e^{2j\frac{\omega_0}{c_0}(x-x_0) + \frac{2\sigma}{c_0}(x-x_0)} u(x-x_0)$$

where C is a constant and $u(\cdot)$ stands for the unit step function. The corresponding ISAR image is the "cloud", spreading through the down-range shown in Fig. 4. This cloud not only makes the image more crowded, but may in some cases obscure some other important scattering centres.

As an example, the finite difference time domain (FDTD) method was used to obtain synthetic time-domain data for the monostatic scattering from the perfectly conducting cavity shown in Fig. 5 [15]. Data was taken for plane waves incident with the electric field polarized in the z-direction. The angle of incidence was varied, in the x-y plane, by 5 degrees over a 180 degree sweep of the target. The ISAR image, constructed using the FDTD data, is shown in Fig. 6. Note that although the outline of the cavity can be discerned, the artifacts due to the resonant nature of the target tend to obscure the image.

5. Physical Optics Approximation Of Scattering

Instead of approaching the problem from the geometrical optics point of view one could formulate the high-frequency scattering problem using physical optics.

5.1. High frequency scattering by a PEC body

Consider a perfectly conducting body illuminated by a plane monochromatic wave, as shown in Fig. 7. The surface of the scatterer is denoted by S and its volume by V. Using the Stratton-Chu formula [16] for the direct integration of the vector electromagnetic wave equation, the magnetic field H_s scattered by the object can be written in terms of the magnetic and electric fields on the surface of the body. Since the boundary conditions are $n \times E = 0$ and $n \cdot H = 0$ the integral equation simplifies to
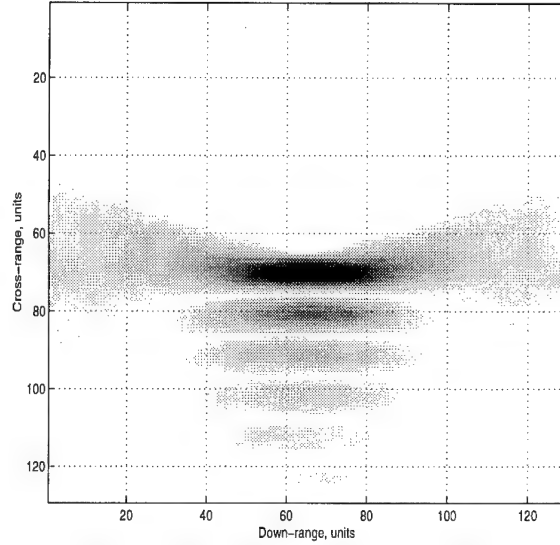
Figure 4. ISAR image of a resonant scatterer.
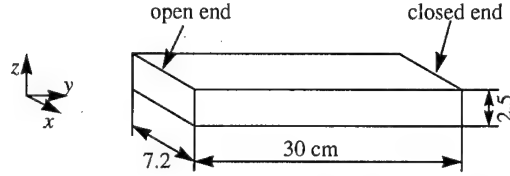
Figure 5. Open ended cavity as a scatterer.
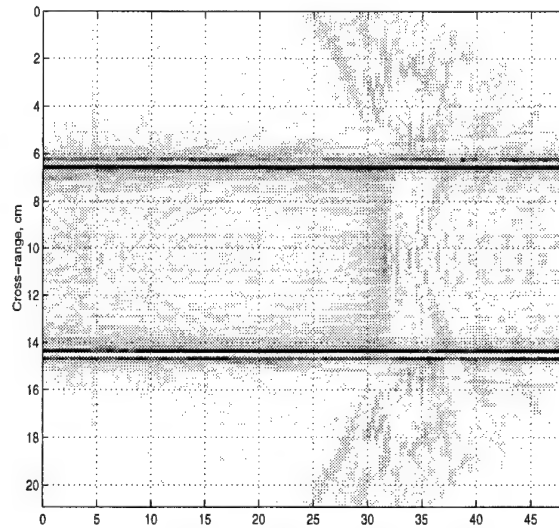
Figure 6. ISAR image of rectangular cavity.

$$H_s = \oint_S [(n \times H) \times \nabla G]ds = -\oint_S \nabla G \times (n \times H)ds \quad (26)$$

which, in turn, can be rewritten for the far-field zone as

$$H_s = -\oint_S \nabla G \times (n \times H)ds = -\oint_S G jk_{3s} \times (n \times H)ds \quad (27)$$

Here n is the unit vector normal to the surface of the scatterer, H is the total magnetic field, G stands for the appropriate Green's function, k_{3i} and k_{3s} are the incident and scattered field wave vectors.
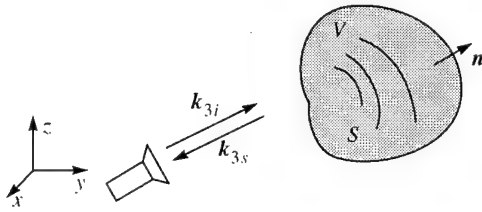

Figure 7. Geometry of the problem

For scatterers that are large compared to the wavelength of the excitation, the physical optics approximation states that the total magnetic field H, tangential to the surface is twice the incident tangential magnetic field on the illuminated side of the scatterer and zero on the shadow side, *i.e.*,

$$n \times H = \begin{cases} 2n \times H_i, & k_{3i} \cdot n < 0 \\ 0, & k_{3i} \cdot n > 0 \end{cases} \quad (28)$$

This approximation allows us to rewrite (27) in the following form

$$\begin{aligned} H_s &= -2j \int_{k_{3i} \cdot n < 0} G k_{3s} \times (n \times H_i) ds \\ &= 2j \int_{k_{3i} \cdot n < 0} G[H_i(k_{3s} \cdot n) - n(k_{3s} \cdot H_i)] ds \end{aligned} \quad (29)$$

If the incident field is taken as a plane wave, then for the monostatic case $k_{3s} = -k_{3i}$ and thus $k_{3i} \cdot H_i = 0$ which reduces (29) to

$$H_s = 2j \int_{k_{3i} \cdot n < 0} G H_i(k_{3s} \cdot n) ds \quad (30)$$

From this equation it can be seen that the polarization of the scattered magnetic field yielded by the physical optics approximation is co-linear with the incident magnetic field polarization. Thus, no cross-polarization information can be obtained from the physical optics approximation. Equation (30) can now be rewritten without loss of generality in terms of the complex spatial vector magnitude of the incident and scattered fields as

$$H_s = 2j \int_{k_{3i} \cdot n < 0} G H_i k_{3s} \cdot ds \quad (31)$$

In the far field, the scattered field can be considered a spherical wave, thus

$$H_s = S(k_{3s}) \frac{e^{j(k_{3s} \cdot x')}}{\sqrt{4\pi x'}} \quad (32)$$

while the incident field is the plane wave

$$H_i = I(k_{3i}) e^{jk_{3i} \cdot x'} \quad (33)$$

Recalling, that the three dimensional Green's function for free space is

$$G = \frac{e^{jk_s \cdot r}}{4\pi r}, \ r = |x - x'| \quad (34)$$

one can rewrite (31) in the form

$$\frac{S(k_{3s})}{I(k_{3i})} = 2j \int_{k_{3i} \cdot n < 0} \frac{e^{jk_{3s} \cdot (r - x')} \sqrt{4\pi} x'}{4\pi r} e^{jk_{3i} \cdot x} k_{3s} \cdot ds \quad (35)$$

In the far field, we have

$$2j \frac{\sqrt{4\pi} x'}{4\pi r} \approx 2j \frac{\sqrt{4\pi} r}{4\pi r} = \frac{j}{\sqrt{\pi}} \text{ and } e^{jk_{3s} \cdot (r - x')} \rightarrow e^{-jk_{3s} \cdot x} \quad (36)$$

thus, (35) can be rewritten, for the monostatic case, as

$$\rho(k_3) = \frac{S(k_3)}{I(k_3)} = \frac{j}{\sqrt{\pi}} \int_{k_3 \cdot n < 0} e^{-2jk_3 \cdot x} k_3 \cdot ds \quad (37)$$

where $k_3 = k_{3s} = -k_{3i}$. Here $\rho(k_3)$ represents the normalized monostatic response.

Let S_k be the surface illuminated by the plane wave propagating in the direction given by the wave vector k_3. Then, assuming a convex object:

$$S = S_k + S_{-k} \quad (38)$$

where S_{-k} is the part of the surface S illuminated by the wave, propagating in the opposite direction to the wave described by wave vector k_3, *i.e.* is the part of the surface illuminated by the wave propagating in the direction described by the wave vector $-k_3$. For this case (37) can be rewritten as

$$\rho(-k_3) = \frac{S(-k_3)}{I(-k_3)} = -\frac{j}{\sqrt{\pi}} \int_{k_3 \cdot n < 0} e^{2jk_3 \cdot x} k_3 \cdot ds \quad (39)$$

or

$$\rho^*(-k_3) = \frac{j}{\sqrt{\pi}} \int_{k_3 \cdot n < 0} e^{-2jk_3 \cdot x} k_3 \cdot ds \quad (40)$$

Adding (37) and (40) yields

$$\begin{aligned} \rho(k_3) + \rho^*(-k_3) &= \frac{j}{\sqrt{\pi}} \int_{k_3 \cdot n > 0} e^{-2jk_3 \cdot x} k_3 \cdot ds + \\ \frac{j}{\sqrt{\pi}} \int_{k_3 \cdot n < 0} e^{-2jk_3 \cdot x} k_3 \cdot ds &= \frac{j}{\sqrt{\pi}} \oint_S e^{-2jk_3 \cdot x} k_3 \cdot ds \end{aligned} \quad (41)$$

According to Gauss theorem, the surface integral can be transformed to a volume integral of the form

$$\rho(k_3) + \rho^*(-k_3) = \frac{j}{\sqrt{\pi}} \oint_S e^{-2jk_3 \cdot x} k_3 \cdot ds$$

$$= \frac{j}{\sqrt{\pi}} \int_V \nabla \cdot (e^{-2jk_3 \cdot x} k_3) dv \qquad (42)$$

$$= \frac{2k_3^2}{\sqrt{\pi}} \int_V e^{-2jk_3 \cdot x} dv$$

where $k_3 = |\dot{k}_3|$.

5.2.   The inverse problem

Equation (42) allows us to solve the direct problem, *i.e.* to find the scattered far field, if the shape of the target is known. A small modification of (42), suggested in [17] allows to solve the inverse problem, *i.e.* to restore the shape of the object if the scattered field is known. This will now be explained.

Let us introduce the following membership function

$$\gamma(x) = \begin{cases} 1, & \forall x \in V \\ 0, & \forall x \notin V \end{cases} \qquad (43)$$

Then the integral (42) can be rewritten as

$$\rho(k_3) + \rho^*(-k_3) = \frac{2k_3^2}{\sqrt{\pi}} \int_V e^{-2jk_3 \cdot x} dv$$

$$= \frac{2k_3^2}{\sqrt{\pi}} \int_{R^3} \gamma(x) e^{-2jk_3 \cdot x} dv \qquad (44)$$

or

$$\sqrt{\pi} \frac{[\rho(k_3) + \rho^*(-k_3)]}{2k_3^2} = \int_{R^3} \gamma(x) e^{-2jk_3 \cdot x} dv \qquad (45)$$

The right part of equation (45) is the three dimensional Fourier transform of the membership function $\gamma(x)$, and can be easily inverted to produce

$$\gamma(x) = \frac{1}{(2\pi)^3} \int_{R^3} \sqrt{\pi} \frac{[\rho(k_3) + \rho^*(-k_3)]}{2k_3^2} e^{-2jk_3 \cdot x} dk_3$$

$$= \frac{\pi^{-5/2}}{16} \int_{R^3} \frac{\rho(k_3)}{k_3^2} e^{-2jk_3 \cdot x} dk_3$$

$$+ \frac{\pi^{-5/2}}{16} \left(\int_{R^3} \frac{\rho(k_3)}{k_3^2} e^{-2jk_3 \cdot x} dk_3 \right)^* \qquad (46)$$

$$= \frac{\pi^{-5/2}}{16} Re \int_{R^3} \frac{\rho(k_3)}{k_3^2} e^{-2jk_3 \cdot x} dk_3$$

Thus, if the response from the target is known for all possible wave vectors k_3, then the membership function $\gamma(x)$ can be exactly reconstructed.

5.3.   Case of incomplete scattering information

Let a scattering information aperture function $A(k_3)$ be defined as unity or zero for all values of k_3 for which the scattered response is known or unknown, respectively, *i.e.*

$$A(k_3) = \begin{cases} 1, & \forall k_3 \text{ for which } \rho(k_3) \text{ is known} \\ 0, & \forall k_3 \text{ for which } \rho(k_3) \text{ is unknown} \end{cases} \qquad (47)$$

and let $a(x)$ be a three-dimensional Fourier transform of $A(k_3)$:

$$a(x) = \frac{1}{(2\pi)^3} \int_{R^3} A(k_3) dk_3 \qquad (48)$$

Then, equation (46) can be rewritten as

$$a(x) * \gamma(x) = \frac{\pi^{-5/2}}{16} Re \int_{R^3} \frac{A(k_3)\rho(k_3)}{k_3^2} e^{-2jk_3 \cdot x} dv \qquad (49)$$

where star denotes 3-D convolution operator. Equation (49) shows that in the case of incomplete angular information the reconstructed ISAR (SAR) image differs from those reconstructed from the complete information. It is 3-D convolution of the "complete" image with some aperture function which depends only on part of the frequency space where measurements are available.

5.4.   Two-dimensional thickness function

If the scattering information is restricted to the two-dimensional (k_x, k_y) plane, then the scattering information aperture must be chosen as the one-dimensional Dirac-delta function in the orthogonal direction to this plane, *i.e.* [18]

$$A(k_3) = \delta(k_z) \qquad (50)$$

Thus, according to (48)

$$a(x) = \frac{1}{(2\pi)^3} \int_{R^3} e^{2jk_3 \cdot x} \delta(k_z) dk_3 = \frac{1}{2\pi} \delta(x)\delta(y) \qquad (51)$$

and the integral equation (49) can be rewritten as

$$T(x, y) = \int_{-\infty}^{\infty} \gamma(x, y, z) dz$$

$$= \int_{-\infty}^{\infty} \frac{\pi^{-5/2}}{16} Re \left(\int_{R^3} \frac{\rho(k_3)}{k_3^2} e^{-2jk_3 \cdot x} dk_3 \right) dz$$

$$= \frac{\pi^{-5/2}}{16} Re \left(\int_{R^3} \frac{\rho(k_3)}{k_3^2} dk_3 \int_{-\infty}^{\infty} e^{-2jk_3 \cdot x} dz \right) \qquad (52)$$

$$= \frac{\pi^{-5/2}}{16} Re \left(\int_{R^3} \frac{\rho(k_3)}{k_3^2} e^{-2jk_2 \cdot x_2} 2\pi \delta(k_z) dk_3 \right)$$

and, finally, taking into account that $k_2 = k_3$ in this case

$$T(x, y) = \frac{\pi^{-3/2}}{8} Re \int_{R^2} \frac{\rho(k_x, k_y)}{k_2^2} e^{-2jk_2 \cdot x_2} dk_2 \qquad (53)$$

It can be easily seen from the equation (52), that the two-dimensional function $T(x, y)$, which can be restored from two-dimensional measurements represents the dimension of the body in the z direction, *i.e.* its thickness [18, 19].

5.5. Conducting plate example

Consider a perfectly conducting rectangular plate of dimension $2a \times 2b \times 2c$ centred on the origin. According to (42) one can write

$$
\begin{aligned}
\rho(\mathbf{k}_3) &= \frac{2|\mathbf{k}_3|^2}{\sqrt{\pi}} \iiint_V e^{-2j\mathbf{k}_3 \cdot \mathbf{x}} dv \\
&= \frac{2|\mathbf{k}_3|^2}{\sqrt{\pi}} \int_{-a}^{a} e^{-2jk_x x} dx \int_{-b}^{b} e^{-2jk_y y} dy \int_{-c}^{c} e^{-2jk_z z} dz \quad (54) \\
&= \frac{2|\mathbf{k}_3|^2}{\sqrt{\pi}} \frac{\sin(2k_x a)}{k_x} \frac{\sin(2k_y b)}{k_y} \frac{\sin(2k_z c)}{k_z}
\end{aligned}
$$

If the measurements are obtained only in the $x-y$ plane than the z component of the $\mathbf{k}_3$ vector is equal to zero: $k_z = 0$ and (54) can be rewritten as

$$
\begin{aligned}
\rho(\mathbf{k}_2) &= \frac{2|\mathbf{k}_3|^2}{\sqrt{\pi}} \frac{\sin(2k_x a)}{k_x} \frac{\sin(2k_y b)}{k_y} 2c \\
&= \frac{4c}{\sqrt{\pi}}(k_x^2 + k_y^2) \frac{\sin(2k_x a)}{k_x} \frac{\sin(2k_y b)}{k_y}
\end{aligned} \quad (55)
$$

The next step is to use the following relation between the components of the $\mathbf{k}_2$ vector and polar coordinates on the $x-y$ plane

$$
k_x = \frac{\omega}{c_0}\cos\varphi, \quad k_y = \frac{\omega}{c_0}\sin\varphi \quad (56)
$$

After plugging (56) into (55) one can find that

$$
\begin{aligned}
\rho(\omega, \varphi) &= \frac{4c}{\sqrt{\pi}}\frac{\omega^2}{c_0^2} \frac{\sin\left(2\frac{\omega}{c_0}a\cos\varphi\right)}{\frac{\omega}{c_0}\cos\varphi} \frac{\sin\left(2\frac{\omega}{c_0}b\sin\varphi\right)}{\frac{\omega}{c_0}\sin\varphi} \\
&= \frac{4c}{\sqrt{\pi}}\frac{1}{\cos\varphi\sin\varphi}\sin(T_a\omega)\sin(T_b\omega) \quad (57) \\
&= \frac{-16c}{\sqrt{\pi}}\frac{1}{\cos\varphi\sin\varphi}[e^{j(T_a+T_b)\omega} + e^{-j(T_a+T_b)\omega} \\
&\quad - e^{j(T_a-T_b)\omega} - e^{-j(T_a-T_b)\omega}]
\end{aligned}
$$

where we have used the following notations

$$
T_a = 2\frac{a\cos\varphi}{c_0}, \quad T_b = 2\frac{b\sin\varphi}{c_0} \quad (58)
$$

If $\varphi = 0$ or $\varphi = \pi$ then the denominator in (57) becomes zero and the true value of the $\rho(\varphi, \omega)$ can be calculated as a limit:

$$
\begin{aligned}
\rho(\pi n, \omega) &= \lim_{\varphi \to \pi n} \rho(\varphi, \omega) \\
&= \lim_{\varphi \to \pi n} \frac{4c}{\sqrt{\pi}} \frac{\sin\left(2\frac{a\cos\varphi}{c_0}\omega\right)\sin\left(2\frac{b\sin\varphi}{c_0}\omega\right)}{\cos\varphi \quad \sin\varphi} \quad (59) \\
&= \frac{4c}{\sqrt{\pi}}\sin\left(\frac{2a}{c_0}\omega\right)\frac{2b}{c_0}\omega
\end{aligned}
$$

and thus, the corresponding time response is

$$
\Re(\pi n, t) = \frac{8bc}{\sqrt{\pi}}\frac{\partial}{\partial t}\left[\delta(t - \frac{2a}{c_0}) - \delta(t + \frac{2a}{c_0})\right] \quad (60)
$$

In the very same manner one can obtain that

$$
\Re(\pm\frac{\pi}{2}, t) = \frac{8ac}{\sqrt{\pi}}\frac{\partial}{\partial t}\left[\delta(t - \frac{2b}{c_0}) - \delta(t + \frac{2b}{c_0})\right] \quad (61)
$$

For any other angles, the inverse Fourier transform, applied to the equation (57), gives

$$
\begin{aligned}
\Re(t, \varphi) &= \frac{16c}{\sqrt{\pi}}\frac{1}{\cos\varphi\sin\varphi} \cdot \\
&\quad [\delta(t - T_1 - T_2) + \delta(t + T_1 + T_2) \quad (62) \\
&\quad -(\delta(t - T_1 + T_2) - \delta(t + T_1 - T_2))]
\end{aligned}
$$

Corresponding ISAR images are shown in Fig. 8-9. It can be clearly seen from the Fig. 8 that scattering from the edges of a plate are dominant, thus producing image similar to "optical" image of the plate, but contaminated by artifacts. At the same time, a narrowband image, shown in Fig. 9 emphasizes point scattering, *i.e.* only sharp changes in target shape can be observed.
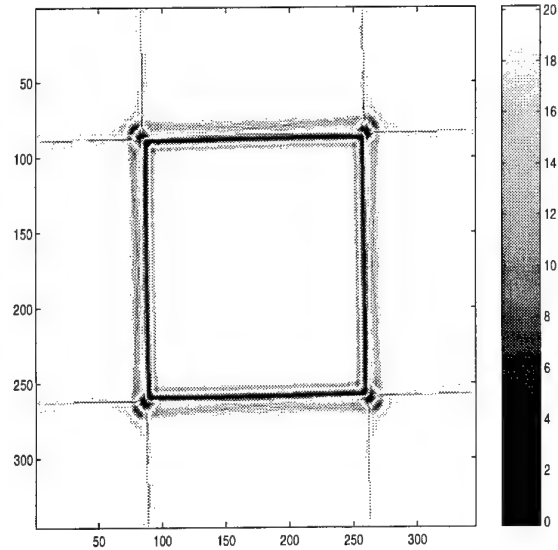

Figure 8. ISAR image of a rectangular plate (PO data).

Yet another important feature can be seen from Fig. 9. Instead of a single point, representing a nearly isotropic point scatterer, two or three different scatterers can be observed. This emphasize the difference between PO and GTD approximations. However, for the case of a rectangular plate it does not introduce a big difference. The results can be quite different if the target has a few closely spaced sharp edges.

6. CONCLUSIONS AND MODERN TRENDS

In this paper we have briefly reviewed the background of ISAR (SAR) microwave imaging and have considered the main methods of image reconstruction from the projections. As we have seen, the ISAR(SAR) image does not coincide with the "true" optical image of the target. Nevertheless, it seems that such an image reflects deep physical properties of the target

under investigation and thus can be used in various applications.

The different techniques considered in this paper are applied to the simplest case of PEC targets in free space. More general setups have been extensively studied during the last decade. One branch of this study is concerned with transparent and semi-trasparent dielectric bodies. It was found that backprojection type algorithms can be used to restore the dielectric profile of a target under test [8] if the dielectric parameters of the body are just slightly different from those of embedding material. If the difference in the electrical parameters is relatively large, direct inversion becomes analytically difficult and a number of iterative techniques have been developed. Most of them require numerical solution of the direct and inverse problem and are time consuming [20]. This, in turn, has stimulated development of multiresolution numerical techniques for direct problems [21]. Another characteristic point of classical microwave reconstruction techniques is the assumption that all the measurements are performed in the far field. However, a large class of the problems (such as land mine detection) requires near-range data acquisition. A few publications can be found in the open literature [22]. Due to redundancy of inverse problems for detection purposes, simplified statistical models have also been developed [23].
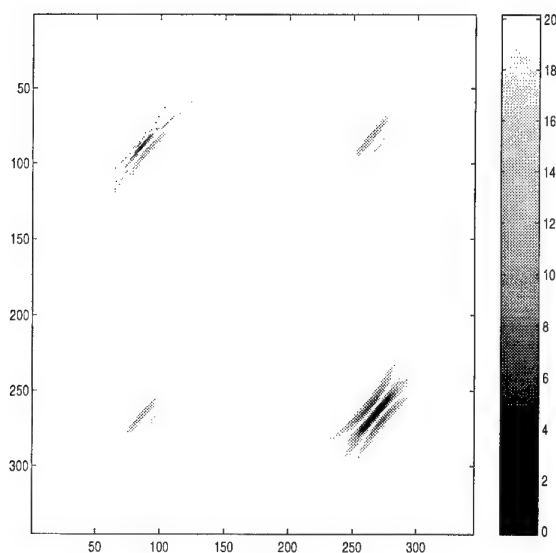
Figure 9. Narrowband ISAR image of a plate (FDTD data)

7. REFERRENCES

[1] V.G. Nebabin, *Methods and Techniques of Radar Recognition*, Artech House, 1995.

[2] D. R. Wehler, *High-Resolution Radar*, Artech House, Boston, 1994.

[3] D. Munson, J.D. O'Brien, W. Jenkins, "A Tomographic Formulation of Spotlight-Mode Synthetic Aperture Radar", *Proc. of the IEEE*, Vol. 71, No. 8, August, 1983, pp. 917-925.

[4] C. Jakowatz, D. Wahl, P. Eichel, D. Ghiglia, and P. Thomson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer, Boston, 1996.

[5] R. M. Mersereau, and A. Oppenheim, "Digital Reconstruction of Multidimensional Signals from Their Projections", *Proceedings of the IEEE*, Vol. 62, No. 10, 1974, pp. 1319-1338.

[6] Z.-H. Cho, J.P. Jones, and M. Singh, *Foundations of Medical Imaging*, John Willey and Sons, New York, 1993.

[7] G. Demoment, "Image Reconstruction and restoration: Overview of Common Estimation Structure and Problems", *IEEE Transaction on Acoustics, Speech and Signal Processing*, Vol. 37, No. 12, December, 1989, pp. 2024-2035.

[8] K.I. Hopcraft, and P.P. Smith, *An Introduction to Electromagnetic Inverse Scattering*, Kluwer, London, 1992.

[9] C. Balanis, *Advanced Engineering Electromagnetics*, John Wiley and Sons, New York, 1989.

[10] L. C. Porter, and R.Moses, "Attributed Scattering Centers for SAR ATR", *IEEE Transactions on Image Processing*, Vol. 6, No. 1, January, 1997, pp. 79-91.

[11] G.T. Herman, *Image Reconstruction from Projections. Implementations and Applications*, Springer-Verlag, New York, 1979.

[12] S. Deans, *The Radon Transform and Some of Its Applications*, John Willey and Sons, New York, 1983.

[13] I.M. Gelphand, and S.G. Gindikin, *Mathematical Problems of Tomography*, American Mathematica Society, 1990.

[14] G. Gaunaurd, and H. Uberal, "Solution of the Inverse Electromagnetic Scattering Problem in the Resonance Case", *IEEE Transaction on Antennas and Propagation*, Vol. 29, No. 2, March, 1981, pp. 293-297.

[15] K.S. Kunz and R.J. Luebbers, *The Finite Difference Time Domain Method for Electromagnetics*, CRC Press, 1993.

[16] J.A. Stratton, *Electromagnetic Theory*, New York: McGraw-Hill, 1941.

[17] N. Bojarski, "A Survay of the Physical Optics Inverse Scattering Identity", *IEEE Transactions on Antennas and Propagation*, Vol. AP-30, No. 5, September, 1982, pp. 980-989.

[18] N. Bojarski, "Low Frequency Inverse Scattering", *IEEE Transactions on Antennas and Propagation*, Vol. AP-30, No. 4, July, 1982, pp. 775-778.

[19] E.J. Rothwell, K.M. Chen, D.P. Nyquist and J.E. Ross, "Time-Domain Imaging of Airborne Targets Using Ultra-Wideband or Short-Pulse Radar," *IEEE Transactions on Antennas and Propagation*, Vol. 43, No. 3, pp.327-329, March 1995.

[20] S.Y. Semenov, "Microwave Tomography: Theoretical and Experimental Investigation of the Iteration Reconstruction Algorithm", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 2, pp.133-141, February, 1998.

[21] M, Krumpholz; LPB. Katehi, "MRTD: new time-domain schemes based on multiresolution analysis", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 44, No. 4, pp.555-71, April, 1996.

[22] A. Broquetas, J. Palau, L. Jofre, and A. Cardama, "Spherical Wave Near-Field Imaging and radar Cross-Section Measurements" *IEEE Transactions on Antennas and Propagation*, Vol. 46, No. 5, pp.730-735, May, 1998.

[23] J. Daba, and M. Bell, "Statistics of Scattering Cross-section of a Small Number of Random Scatterers", *IEEE Transaction on Antennas and Propagation*, Vol. 43, No. 8, pp. 773-783, August, 1995.

Two-Dimensional Inverse Profiling: Nonlinear Optimization and Embedding

Anton G. Tijhuis, Kamal Belkebir and Amélie Litman
Faculty of Electrical Engineering
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Jean-Michel Geffrin and Jean-Charles Bolomey
Lab. des Signaux et Systèmes
CNRS/SUPELEC
Plateau du Moulon
91192 Gif-sur-Yvette Cedex
France

1. SUMMARY

Previous work by the first two authors has shown that iterative procedures for solving nonlinear inverse-scattering problems that are based on a local linearization can be used to reconstruct objects of large contrast with a given resolution, provided that multiple-frequency information is available. This approach has become feasible since combining the CGFFT method with a special extrapolation procedure leads to an extremely fast solution of the corresponding direct-scattering problem for varying angle of incidence and contrast. Presently, we are trying to overcome two of the limitations of this approach. First, we replace the "update step" by a line search in a nonlinear optimization procedure. Second, we use the scattering response of the estimated object in a homogeneous environment to reconstruct objects in more realistic measurement set ups like a water-filled metal cylinder.

2. FORMULATION OF THE PROBLEM

We consider an inhomogeneous, lossless dielectric cylinder in an observation domain $\mathcal{D}_O$, embedded in vacuum and excited by a time-harmonic electric line source on a circular contour $\partial \mathcal{D}_O$ around the cylinder (Fig. 1). "Complete" scattering data are available, i.e. the scattered electric field is known on $\partial \mathcal{D}_O$ for a line source anywhere on that contour. The question is to what extent we can reconstruct the relative permittivity $\varepsilon_r(x, y)$ from this field information. A Fourier transformation with respect to the angular coordinate shows that: (i) any incident field in $\mathcal{D}_O$ can be generated by a surface current on $\partial \mathcal{D}_O$; (ii) any scattered field in $\bar{\mathcal{D}}_O$ is completely determined by its value on $\partial \mathcal{D}_O$. This implies that excitation or detection in $\bar{\mathcal{D}}_0$ does not give new information, which explains the designation "complete".

Mathematically, the field caused by each line source can be identified as a Green's function, i.e., the solution of the second-order differential equation

$$[\nabla_T^2 - s^2 \varepsilon_r(\boldsymbol{\rho})] G(\boldsymbol{\rho}, \boldsymbol{\rho}_S) = -\delta(\boldsymbol{\rho} - \boldsymbol{\rho}_S) \qquad (1)$$

that satisfies the radiation condition as $\rho = |\boldsymbol{\rho}| \to \infty$. In (1), s is a complex frequency with $\mathrm{Re}(s) \geq 0$.

For the function $G(\boldsymbol{\rho}, \boldsymbol{\rho}_S)$ the following contrast-source
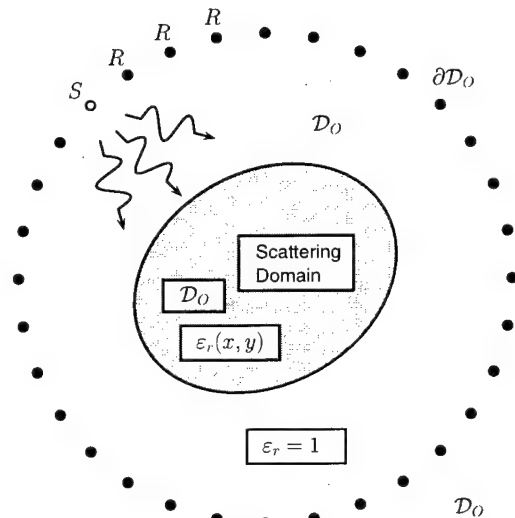
FIGURE 1: typical inverse-scattering geometry.

integral relation can be derived:

$$G(\boldsymbol{\rho}, \boldsymbol{\rho}_S) = G^{ref}(\boldsymbol{\rho}, \boldsymbol{\rho}_S) - s^2 \iint_{\mathcal{D}_O} G^{ref}(\boldsymbol{\rho}, \boldsymbol{\rho}') \times$$
$$[\varepsilon_r(\boldsymbol{\rho}') - \varepsilon_r^{ref}(\boldsymbol{\rho}')] G(\boldsymbol{\rho}', \boldsymbol{\rho}_S) \, dA', \quad (2)$$

which holds for any $\{\boldsymbol{\rho}, \boldsymbol{\rho}_S\} \in \mathcal{D}_\infty$. In (2), the superscript ^{ref} pertains to a known reference configuration.

2. FORWARD PROBLEM

Before we address the inverse problem, we must be able to determine the Green's function $G(\boldsymbol{\rho}, \boldsymbol{\rho}_S)$ for a known permittivity $\varepsilon_r(\boldsymbol{\rho})$. This field is obtained from the so-called *field equation*, which is obtained from (2) by taking $\boldsymbol{\rho} \in \mathcal{D}_O$ and $\varepsilon_r^{ref} = 1$. We obtain the contrast-source integral equation:

$$G(\boldsymbol{\rho}, \boldsymbol{\rho}_S) = G_0(\boldsymbol{\rho}, \boldsymbol{\rho}_S) - s^2 \iint_{\mathcal{D}_O} G_0(\boldsymbol{\rho}, \boldsymbol{\rho}') \times$$
$$[\varepsilon_r(\boldsymbol{\rho}') - 1] G(\boldsymbol{\rho}', \boldsymbol{\rho}_S) \, dA'. \quad (3)$$

In (3), $G_0(\boldsymbol{\rho}, \boldsymbol{\rho}')$ is the two-dimensional Green's function of free space

$$G_0(\boldsymbol{\rho}, \boldsymbol{\rho}') = \frac{1}{2\pi} K_0(s|\boldsymbol{\rho} - \boldsymbol{\rho}'|), \qquad (4)$$

where K_0 is the modified Bessel function of the second kind of order 0.

Equation (3) is discretized as described in [2]. First, we decompose $G_0(\boldsymbol{\rho}, \boldsymbol{\rho}')$ into a stationary part, which is logarithmically singular as $R = |\boldsymbol{\rho} - \boldsymbol{\rho}'| \to 0$, and a part which, apart from an additive constant, behaves as $R^2 \ln(R)$. This results in two integrals, which can both be approximated with the aid of piecewise-bilinear interpolation with mesh size h. This results in a convolution-type form that is accurate up to $\mathcal{O}[h^2 \ln(h)]$ as $h \to 0$. We choose h so small that the error in determining $G(\boldsymbol{\rho}, \boldsymbol{\rho}_S)$ does not affect the inversion procedure. The resulting discretized system is solved by CGFFT, using the extrapolation procedure outlined in [1] to generate the successive initial estimates. This procedure is known as "marching on in anything"; in the present application, we march either in angle of incidence or in contrast. Typically, the full scattering matrix is available at the cost of at most a few field computations with the original CGFFT procedure.

3. COST FUNCTION AND LINEARIZATION

In numerical schemes for solving the inverse-scattering problem, it is convenient to introduce a *parameterization* of the unknown permittivity:

$$\tilde{\varepsilon}(\boldsymbol{\rho}) = \sum_\alpha \tilde{\varepsilon}_\alpha \, \phi_\alpha(\boldsymbol{\rho}) \qquad (5)$$

where $\phi_\alpha(\boldsymbol{\rho})$ is a known expansion function. Within the space spanned by these expansion functions, we define the "reconstruction" as the linear combination for which the squared error

$$ERR = \int_0^{2\pi} d\varphi_R \int_0^{2\pi} d\varphi_S \left| \tilde{G}(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S) - G(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S) \right|^2, \quad (6)$$

is minimized. In (6), $G(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S)$ is the known field at receiver position $\boldsymbol{\rho} = \boldsymbol{\rho}_R$ for a source at $\boldsymbol{\rho} = \boldsymbol{\rho}_S$, and $\tilde{G}(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S)$ is the field in the configuration corresponding to the set of parameters $\{\tilde{\varepsilon}_\alpha\}$. With (5) and (6), the inverse-scattering problem has been reduced to an optimization problem for a fixed number of parameters.

Most inversion schemes are based on a linearization of the cost function (6) around some "best" estimate $\bar{\varepsilon}(\boldsymbol{\rho}) = \sum_\alpha \bar{\varepsilon}_\alpha \phi_\alpha(\boldsymbol{\rho})$. Taking $\varepsilon^{ref} = \bar{\varepsilon}$ results in the integral relation:

$$G(\boldsymbol{\rho}, \boldsymbol{\rho}') = \bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}') - s^2 \iint_{\mathcal{D}_O} [\varepsilon_r(\boldsymbol{\rho}'') - \bar{\varepsilon}_r(\boldsymbol{\rho}'')] \times$$
$$\bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}'') \, G(\boldsymbol{\rho}'', \boldsymbol{\rho}') \, dA, \quad (7)$$

for $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathcal{D}_\infty$. We use the same representation for $G(\boldsymbol{\rho}'', \boldsymbol{\rho}')$ in the integrand on the right-hand side and use the reciprocity relation $\bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}') = \bar{G}(\boldsymbol{\rho}', \boldsymbol{\rho})$ to obtain the *approximate equation*:

$$s^2 \iint_{\mathcal{D}_O} [\varepsilon_r(\boldsymbol{\rho}) - \bar{\varepsilon}_r(\boldsymbol{\rho})] \, \bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}_R) \, \bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}_S) \, dA$$
$$= \bar{G}(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S) - G(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S) + \mathcal{O}([\varepsilon_r - \bar{\varepsilon}_r]^2). \quad (8)$$

In this equation, we replace $\varepsilon(\boldsymbol{\rho})$ by the parameterized version (5), and we minimize the cost function (6), augmented by a regularization term that favours a smooth solution. This should result in a "better" estimate $\tilde{\varepsilon}(\boldsymbol{\rho})$.

Linearized inversion schemes like distorted-wave Born and Newton-Kantorovitch essentially amount to alternately determining $\bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}_S)$ by the procedure outlined in Section 2, and $\tilde{\varepsilon}(\boldsymbol{\rho})$ as explained above. This procedure gives rise to two fundamental questions.

• First, how much information can be retrieved? To answer this question, we consider a small perturbation (a "pixel") in a known configuration, and we use the distorted-wave Born approximation given in (8) to reconstruct it in a single step. This step is representative of the final stage of the iterative procedure. We formally arrive at a SAR-like processing scheme which shows that the resolution improves with increasing $\omega \sqrt{\bar{\varepsilon}_r(\boldsymbol{\rho})}$. This has been confirmed by implementing the reconstruction scheme for the special case where the estimated configuration is a homogeneous circular cylinder.

• Second, will the scheme converge? The dynamic range of the procedure depends on the error in the linearization, i.e., on the magnitude of the term of $\mathcal{O}([\varepsilon_r - \bar{\varepsilon}_r]^2)$ in (8). For a general configuration, it is hard to give a quantitative estimate of this "Born error". Therefore, we have investigated its behavior from closed-form expressions for homogeneous circular cylinders. It turns out that the "Born error" increases with increasing contrast $|\varepsilon_r - \bar{\varepsilon}|$, difference in object dimension, and frequency ω.

To reconcile these observations, we start with a small ω to obtain the "trend". In a few steps, we then increase the frequency to obtain more and more "detail", using previous results as initial estimates for $\varepsilon_r(\boldsymbol{\rho})$.

3. EXTENSIONS

With the fast forward solution of Section 2 and the theoretical analysis of Section 3, linearized inversion schemes have again become attractive. However, the availability of the full scattering matrix also offers the possibility to overcome some of the limitations of the scheme outlined in Section 3.

• A first idea is to replace the determination of $\tilde{\varepsilon}(\boldsymbol{\rho})$ by a line search in a nonlinear optimization scheme. Substituting the linearized equation in the cost function gives the *profile gradient* in closed form.

$$\int_0^{2\pi} \int_0^{2\pi} |\tilde{G}(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S) - G(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S)|^2 d\varphi_R \, d\varphi_S$$
$$\approx \int_0^{2\pi} \int_0^{2\pi} |\bar{G}(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S) - G(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S)|^2 d\varphi_R \, d\varphi_S \quad (9)$$
$$+ 2\,\text{Re} \left\{ s^2 \iint_{\mathcal{D}_O} [\tilde{\varepsilon}(\boldsymbol{\rho}) - \bar{\varepsilon}(\boldsymbol{\rho})] \int_0^{2\pi} \int_0^{2\pi} \bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}_R) \times \right.$$
$$\left. \bar{G}(\boldsymbol{\rho}, \boldsymbol{\rho}_S) [\bar{G}(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S) - G(\boldsymbol{\rho}_R, \boldsymbol{\rho}_S)]^* \, d\varphi_R \, d\varphi_S \, dA \right\}$$

By "marching on in search parameter" we obtain a similar efficiency of the successive field computations as in the linearized schemes. This idea has recently been implemented in combination with the BFGS method [3].

• A second idea is to consider a more general environment. The CGFFT method can only be applied to objects in a homogeneous environment. Practical measurements, e.g. for biological tissue, must be carried out in a more complicated set-up like a water-filled metal cylinder. Such

a set-up is presently being developed by the last two authors (Fig. 2). In that case, a direct computation of the field is relatively complicated from a computational point of view. Hence, the idea is to use the procedure outlined above to determine the complete scattering response of the estimated object in a homogeneous environment, and subsequently use a spectral formulation to "embed" this object in the actual environment [4]. That environment is then characterized once by experimental and/or computational techniques.
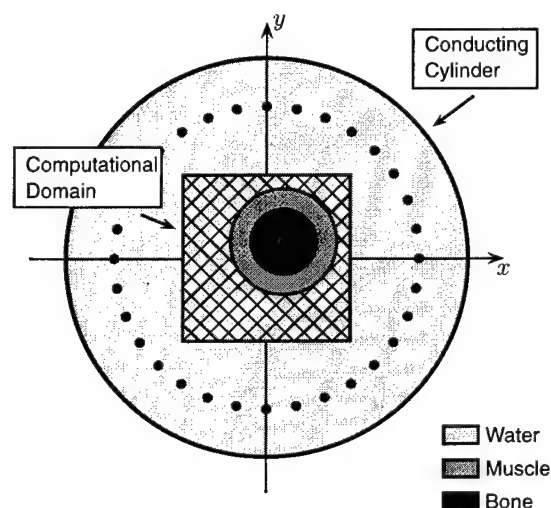
FIGURE 2: set-up for biological tissue.

At the conference, both ideas will be discussed, and the first numerical results will be presented. To this end, the modified algorithm will be applied to theoretical as well as experimental results.

REFERENCES
1. Tijhuis, A.G. and Peng, Z.Q., "Marching-on-in-frequency method for solving integral equations in transient electromagnetic scattering," IEE Proc. H, vol. 138, no. 4, August 1991, pp. 347–355.
2. Peng, Z.Q. and Tijhuis, A.G., "Transient scattering by a lossy dielectric cylinder: marching-on-in-frequency approach," JEWA, vol. 7, no. 4, August 1993, pp. 739–763.
3. Fletcher, R., "Practical Methods of Optimization", 2nd edition, Wiley, Chichester, 1990.
4. Fokkema, J.T. and Van den Berg, P.M., "Seismic Applications of Acoustic Reciprocity", Elsevier, Amsterdam, 1993.

Non-linear Inversion Based on Contrast Source Gradients

Peter M van den Berg, Richard F Bloemenkamp
Laboratory of Electromagnetic Research
Faculty of Electrical Engineering
Centre for Technical Geoscience
Delft University of Technology
P.O. Box 5031
2600 GA Delft
The Netherlands

A.P.M. Zwamborn
TNO Physics and Electronics Laboratory
Oude Waalsdorperweg 63
P.O. Box 96864
2509 JG 's Gravenhage
The Netherlands

1. ABSTRACT

This paper describes a simple algorithm for reconstructing the complex index of refraction of a bounded object immersed in a known background from a knowledge of how the object scatters known incident radiation. The method described here is versatile accommodating both spatially and frequency varying incident fields and allowing *a priori* information about the scatterer to be introduced in a simple fashion. Numerical results show that this new algorithm outperforms the modified gradient approach which until now has been one of the most effective reconstruction algorithms available.

2. INTRODUCTION

The problem of reconstructing the complex index of refraction of a bounded object immersed in a known background medium, from a knowledge of how the object scatters known incident acoustic or electromagnetic radiation, has received a tremendous amount of attention in the past decade. Almost all reconstruction algorithms rely in some way upon the Lippmann-Schwinger equation or domain integral equation for the field inside the scattering object as well as the related integral representation for the field outside the object.

The present paper describes a simple algorithm for reconstructing unknown contrasts which is extremely versatile, accommodating both spatially and frequency varying incident fields and allowing for the introduction of *a priori* information, such as positivity constraints, in a simple fashion. The algorithm is a variant of the source type integral equation (STIE) method introduced by Habashy *et al.* [11] on one hand and the modified gradient approach used by the authors in [12], [13], [14] on the other. Numerical results show that, despite the simplicity of the algorithm, it outperforms the modified gradient approach which has been one of the most effective reconstruction algorithms available until now [25]. We present here the simplest version of the algorithm wherein we treat scalar waves in $\mathbf{R}^2$ for bodies immersed in a homogeneous background.

3. NOTATION AND PROBLEM STATEMENT

Denote by p and q position vectors in $\mathbf{R}^2$ and let B denote a bounded, not necessarily connected, scattering object (or objects) whose location and index of refraction or contrast is unknown but which is known to lie within another, larger, bounded simply connected domain D. If $u_j^{\text{inc}}(p) = u^{\text{inc}}(p, q_j, k_j)$ denotes an incident wave with wavenumber k_j (assumed to be real) and source point q_j (q_j is replaced by the unit vector $\hat{q}_j$ for plane waves) then for a large class of scattering problems the total field in D is known to satisfy the integral equation

$$
\begin{aligned}
u_j(p) &= u_j^{\text{inc}}(p) \\
&+ k_j^2 \int_D G_j(p, q)\, \chi_j(q)\, u_j(q)\, \mathrm{dv}(q)\,, \quad (1)
\end{aligned}
$$

where $G_j(p, q)$ denotes the Green function of the background medium,

$$
G_j(p, q) = \frac{\mathrm{i}}{4} H_0^{(1)}(k_j|p - q|)\,, \quad (2)
$$

and

$$
\chi_j = \frac{k^2(q, k_j)}{k_j^2} - 1\,. \quad (3)
$$

We assume that, while the spatial dependence of χ_j is unknown, the frequency dependence is known, so that, for example, in electromagnetics for a Maxwell medium,

$$
k_j = \omega_j(\varepsilon_0\mu_0)^{\frac{1}{2}} \quad (4)
$$

and

$$
\chi_j(q) = \frac{\varepsilon(q) - \varepsilon_0}{\varepsilon_0} + \mathrm{i}\frac{\sigma(q)}{\omega_j\varepsilon_0}\,, \quad (5)
$$

where ε_0 and μ_0 are the permittivity and the permeability of the (lossless) background, while $\varepsilon(q)$ and $\sigma(q)$ are the permittivity and conductivity of the scatterer which is assumed to be nonmagnetic. Equation (5) may also be written as

$$
\chi_j(q) = \chi^r(q) + \mathrm{i}\chi^i(q)\eta_j, \quad \eta_j = \frac{\omega_0}{\omega_j} = \frac{k_0}{k_j}\,, \quad (6)
$$

where k_0 is the wavenumber for $j = 0$. This equation simplifies to

$$
\chi_j(q) := \chi(q) = \chi^r(q) + \mathrm{i}\chi^i(q)\,, \quad (7)
$$

if all the measurements are made at the same frequency, say at $k_j = k_0$.

Observe that if p is not in B then χ_j vanishes, but if the location of B is unknown then it is not known *a priori* where χ_j vanishes. However with the assumption that $B \subset D$ it is known that χ_j vanishes for p outside D. In fact denoting by S a domain (or curve, or a discrete collection of points) outside of D where the scattered field is measured to be $f_j(p)$, (1) becomes

$$f_j(p) = k_j^2 \int_D G_j(p,q) \, \chi_j(q) \, u_j(q) \, \mathrm{dv}(q) , \qquad (8)$$
$$p \in S ,$$

if there is no noise or error in the measurements. But error free data are extremely unlikely and we do not assume that (8) holds exactly. Rewriting (1) and (8) in symbolic form we have the object or state equations

$$u_j = u_j^{\mathrm{inc}} + G_j^D \chi_j u_j , \quad p \in D , \qquad (9)$$

and the data equations

$$f_j = G_j^S \chi_j u_j , \quad p \in S , \qquad (10)$$

the superscripts D and S on the operators defined implicitly in (1) and (8) are added to accentuate the location of the point p, since the operators are identical in all other respects.

4. THEORETICAL BACKGROUND

In the absence of other *a priori* information, (9) and (10) are the only equations we have relating the unknown contrast χ_j (which, recall, consists of at most two unknown real valued functions) and the unknown fields u_j in D. The known data consist of the incident fields, u_j^{inc}, the wavenumbers, k_j, the measured data, f_j, and the test domain D. Equations (9) and (10) are linear in each of the unknowns χ_j and u_j, but since both are unknown the problem is in fact mildly nonlinear. Of course the dependence of the fields u_j on the contrast χ_j is highly non-linear. This may be seen by writing the formal inverse of (10) as

$$u_j = (I - G_j^D \chi_j)^{-1} u_j^{\mathrm{inc}} . \qquad (11)$$

This form has been utilized in a number of inversion methods. Introducing it into the data equations we obtain

$$f_j = G_j^S [\chi_j (I - G_j^D \chi_j)^{-1} u_j^{\mathrm{inc}}] , \qquad (12)$$

wherein the non-linearity of the inverse problem is clearly exposed. Approximating the inverse operator by

$$(I - G_j^D \chi_j)^{-1} \approx I , \qquad (13)$$

leads to the Born approximation, while in iterative methods, where a sequence $\{\chi_{j,n}\}$ is constructed, the approximation

$$(I - G_j^D \chi_{j,n})^{-1} \approx (I - G_j^D \chi_{j,n-1})^{-1} \qquad (14)$$

gives rise to the iterative Born method [20], while the linearization of

$$(I - G_j^D \chi_{j,n})^{-1} \approx$$
$$[I - G_j^D \chi_{j,n-1} - G_j^D (\chi_{j,n} - \chi_{j,n-1})]^{-1} \qquad (15)$$

in terms of $\Delta\chi_{j,n} = \chi_{j,n} - \chi_{j,n-1}$, namely,

$$(I - G_j^D \chi_{j,n})^{-1} \approx$$
$$[I + (I - G_j^D \chi_{j,n-1})^{-1} G_j^D \Delta\chi_{j,n}](I - G_j^D \chi_{j,n-1})^{-1} \qquad (16)$$

leads to the Newton-Kantorovich method [17], [19] which has been shown to be equivalent to the Distorted Born approach [6]. Observe that at each step in the iteration it is necessary to compute the action of the operator $[I - G_j^D \chi_{j,n-1}]^{-1}$ for known $\chi_{j,n-1}$ This means that forward problems, or direct scattering problems must be solved at each iterative step.

A method which avoids the necessity of solving forward problems completely was proposed by the authors [12] and was refined [13], [14], [25] and extended [24]. This modified gradient method involves the simultaneous construction of sequences $\{u_{j,n}\}$ and $\{\chi_{j,n}\}$ to minimize the error in a cost functional consisting of the normalized errors in both state equations (9) and data equations (10). It has proven to be very effective in a large number of numerical tests using both synthetic and experimental data.

Because the contrast and fields occur as a product, many workers have introduced the quantity

$$w_j = \chi_j u_j , \qquad (17)$$

which is called a contrast source since u_j satisfies the equation

$$(\nabla^2 + k_j^2)u_j = -k_j^2 w_j \quad \text{in } B , \qquad (18)$$

Then the data equations become

$$f_j = G_j^S w_j , \qquad (19)$$

while the state equations become

$$u_j = u_j^{\mathrm{inc}} + G_j^D w_j , \qquad (20)$$

or, with (17),

$$\chi u_j^{\mathrm{inc}} = w_j - \chi_j G_j^D w_j . \qquad (21)$$

Equation (19) is called by some a source type integral equation and it has a long history. It is a classic ill-posed equation and for a time there was considerable attention paid to the question of uniqueness, e.g., [1], [8], [2]. It was shown that there exist non-trivial solutions of the homogeneous form of (19), although it was argued by some that uniqueness could be restored from physical considerations. A good summary of the debate is given by Devaney and Sherman [9] and the responses by Bojarski [3] and Stone [18]. It is not our intent to renew this controversy since it is now well accepted that non-trivial solutions of (19) exist. Moreover it has also been shown that the minimum norm solution of (19), the solution produced, for example, by the conjugate gradient method, is not the appropriate physical solution. Nonetheless this source type equation has served as an essential ingredient in many inversion procedures, e.g., [10], [4], [5], [7], [11]. Habashy *et al.* [11] present an inversion method wherein the minimum norm solution of (19) is found first and then a basis for the orthogonal complement of this solution is constructed in terms of which the physical solution is sought to satisfy (21). Van den Berg and Haak [22] proposed a variant of this technique wherein the full minimum norm solution is not found but rather it is sought iteratively, using conjugate gradient steps, with the contrast updated at

each step to satisfy (21) and a new source defined through (17). This approach yielded promising numerical results however the error did not decrease monotonically.

In the present paper we propose a method which combines spirit of the approach of Van den Berg and Haak [22] using the source type integral equation with that of the modified gradient approach by seeking linear updates in the source in an error reducing method which does not require the solution of any forward problem.

5. CONTRAST SOURCE INVERSION METHOD

As in the modified gradient as well as the Van den Berg-Haak approach we simultaneously construct sequences of sources $w_{j,n}$, fields $u_{j,n}$ and contrasts $\chi_{j,n}$ to minimize a cost functional. Rather than choosing a cost functional consisting only of errors in the data equation, as Van den Berg and Haak did, we define the cost functional

$$
\begin{aligned}
F &= \frac{\sum_j \|f_j - G_j^S w_j\|_S^2}{\sum_j \|f_j\|_S^2} \\
&+ \frac{\sum_j \|\chi_j u_j^{\text{inc}} - w_j + \chi_j G_j^D w_j\|_D^2}{\sum_j \|\chi_j u_j^{\text{inc}}\|_D^2} , \quad (22)
\end{aligned}
$$

where $\| \cdot \|_S^2$ and $\| \cdot \|_D^2$ denote the norms on $L_2(S)$ and $L_2(D)$, respectively. The normalization is chosen so that both terms are equal to one if $w_j = 0$. The first term measures the error in the data equations and the second term measures the error in the form of the state equations given in (21). This is a quadratic functional in w_j, but highly nonlinear in χ_j. We propose an iterative minimization of this cost functional using an alternating method which first updates w_j and then updates χ_j. Thus we construct sequences $\{w_{j,n}\}$ and $\{\chi_{j,n}\}$, for $n = 0, 1, 2, \cdots$, in the following manner.

Define the data error to be

$$
\rho_{j,n} = f_{j,n} - G_j^S w_{j,n} , \quad (23)
$$

and the state error to be

$$
r_{j,n} = \chi_{j,n} u_{j,n} - w_{j,n} , \quad (24)
$$

where

$$
u_{j,n} = u_j^{\text{inc}} + G_j^D w_{j,n} . \quad (25)
$$

Now suppose $w_{j,n-1}$ and $\chi_{j,n-1}$ are known. We update w_j by

$$
w_{j,n} = w_{j,n-1} + \alpha_{j,n} v_{j,n} , \quad (26)
$$

where $\alpha_{j,n}$ is constant and the update directions $v_{j,n}$ are functions of position.

The update directions are chosen to be the Polak-Ribière conjugate gradient directions

$$
\begin{aligned}
v_{j,0} &= 0, \\
v_{j,n} &= g_{j,n} + \frac{\langle g_{j,n}, g_{j,n} - g_{j,n-1} \rangle_D}{\langle g_{j,n-1}, g_{j,n-1} \rangle_D} v_{j,n-1} , \quad (27) \\
& \qquad n \geq 1 ,
\end{aligned}
$$

where $g_{j,n}$ is the gradient (Frechet derivative) of the cost functional with respect to w_j evaluated at $w_{j,n-1}, \chi_{j,n-1}$, while $\langle \cdot, \cdot \rangle_D$ denotes the inner product on $L_2(D)$. Explicitly this found to be

$$
g_{j,n} = -\frac{G_j^{S^*} \rho_{j,n-1}}{\sum_k \|f_k\|_S^2} - \frac{r_{j,n-1} - G_j^{S^*}(\overline{\chi}_{j,n-1} r_{j,n-1})}{\sum_k \|\chi_{k,n-1} u_k^{\text{inc}}\|_D^2} , \quad (28)
$$

where $G_j^{S^*}$ and $G_j^{D^*}$ are the adjoints of G_j^S and G_j^D mapping $L_2(S)$ into $L_2(D)$ and $L_2(D)$ into $L_2(D)$, respectively. Further the overbar denotes complex conjugate.

With the update directions completely specified the constant $\alpha_{j,n}$ is determined to minimize the cost functional

$$
\begin{aligned}
F &= \frac{\sum_j \|f_j - G_j^S w_{j,n}\|_S^2}{\sum_j \|f_j\|_S^2} \\
&+ \frac{\sum_j \|\chi_{j,n-1} u_{j,n} - w_{j,n}\|_D^2}{\sum_j \|\chi_{j,n-1} u_j^{\text{inc}}\|_D^2} \\
&= \frac{\sum_j \|\rho_{j,n-1} - \alpha_{j,n} G_j^S v_{j,n}\|_S^2}{\sum_j \|f_j\|_S^2} \\
&+ \frac{\sum_j \|r_{j,n-1} - \alpha_{j,n}(v_{j,n} - \chi_{j,n-1} G_j^D v_{j,n})\|_D^2}{\sum_j \|\chi_{j,n-1} u_j^{\text{inc}}\|_D^2} \quad (29)
\end{aligned}
$$

and is found explicitly to be

$$
\begin{aligned}
\alpha_{j,n} &= \frac{a+b}{c+d} , \quad (30) \\
a &= \frac{\langle \rho_{j,n-1}, G_j^S v_{j,n} \rangle_S}{\sum_k \|f_k\|_S^2} , \\
b &= \frac{\langle r_{j,n-1}, v_{j,n} - \chi_{j,n-1} G_j^D v_{j,n} \rangle_D}{\sum_k \|\chi_{k,n-1} u_k^{\text{inc}}\|_D^2} , \\
c &= \frac{\|G_j^S v_{j,n}\|_S^2}{\sum_k \|f_k\|_S^2} , \\
d &= \frac{\|v_{j,n} - \chi_{j,n-1} G_j^D v_{j,n}\|_D^2}{\sum_k \|\chi_{k,n-1} u_k^{\text{inc}}\|_D^2} ,
\end{aligned}
$$

where $\langle \cdot, \cdot \rangle_S$ denotes the inner product on $L_2(S)$.

Once $w_{j,n}$ is determined, $u_{j,n}$ is obtained via (25) and (26) as

$$
u_{j,n} = u_{j,n-1} + \alpha_{j,n} G_j^D v_{j,n}, \quad (31)
$$

and we then seek χ_j to minimize the cost functional

$$
F_D = \frac{\sum_j \|\chi_j u_{j,n} - w_{j,n}\|_D^2}{\sum_j \|\chi_j u_{j,n}\|_D^2} . \quad (32)
$$

Since this minimization is not so easy especially in the case of *a priori* information, we use a minimization in two steps. First we minimize the much simpler cost functional

$$
F_D' = \sum_j \|\chi_j u_{j,n} - w_{j,n}\|_D^2 , \quad (33)
$$

and then we use the found contrast as an optimization direction in a line minimization to minimize (32). This two-step minimization technique guarantees that the process is always error reducing and allows for easy implementation of *a priori* information or constraints on χ_j. Since we finally need to minimize the cost functional of (32), we define the contrast function that minimizes (33), $\tilde{\chi}_j$, through

$$
F_D'(\tilde{\chi}_j) = \min(F_D') \quad (34)
$$

Restricting attention to inhomogeneities complying with the Maxwell model, (6), and in the absence of any a priori information on χ, we find that, [23], F_D' is minimized by choosing

$$
\begin{aligned}
\tilde{\chi}_n^r &= \frac{\sum_j \text{Re}(w_{j,n} \overline{u}_{j,n})}{\sum_j |u_{j,n}|^2} , \\
\tilde{\chi}_n^i &= \frac{\sum_j \eta_j \text{Im}(w_{j,n} \overline{u}_{j,n})}{\sum_j \eta_j^2 |u_{j,n}|^2} . \quad (35)
\end{aligned}
$$

However if a priori information is available then it is relatively simple to incorporate it in choosing $\tilde{\chi}_n$. If either χ^r or χ^i is known, then we merely use this known value in place of either the first equality or second equality of (35). Thus, for example, if $\chi^i = 0$, we limit our reconstruction procedure to $\tilde{\chi}_n^r$ from the outset. If we have a priori information that χ^r and χ^i are positive, we use, [23]

$$\tilde{\chi}_n^r = \left\{ \frac{\sum_j \left(\frac{\text{Re}(w_{j,n}\overline{u}_{j,n})}{|u_{j,n}|} \right)^2}{\sum_j |u_{j,n}|^2} \right\}^{\frac{1}{2}}, \qquad (36)$$

$$\tilde{\chi}_n^i = \left\{ \frac{\sum_j \left(\frac{\text{Im}(w_{j,n}\overline{u}_{j,n})}{|u_{j,n}|} \right)^2}{\sum_j \eta_j^2 |u_{j,n}|^2} \right\}^{\frac{1}{2}}. \qquad (37)$$

These choices of $\tilde{\chi}_n^r$ and $\tilde{\chi}_n^i$ coincide with those obtained by Kohn and McKenney [15] for an optimization problem with a positivity constraint and employed in the modified gradient algorithm [14] with good results.

Next a line minimization is used to make the cost functional of equation (32) error reducing. We introduce a contrast update direction as

$$d_n = \tilde{\chi}_n - \chi_{n-1} \qquad (38)$$

and we write χ_n as

$$\chi_n = \chi_{n-1} + \theta d_n \qquad (39)$$

Then θ is chosen to minimize the cost functional of equation (32)

$$\frac{\sum_j \|\chi_n u_{j,n} - w_{j,n}\|_D^2}{\sum_j \|\chi_n u_j^{\text{inc}}\|_D^2}$$

$$= \frac{\sum_j \|\chi_{n-1} u_{j,n} - w_{j,n} + \theta(\tilde{\chi}_n - \chi_{n-1})u_{j,n}\|_D^2}{\sum_j \|\chi_{n-1} u_j^{\text{inc}} + \theta(\tilde{\chi}_n - \chi_{n-1})u_j^{\text{inc}}\|_D^2}$$

$$= \frac{a\,\theta^2 + 2b\,\theta + c}{A\,\theta^2 + 2B\,\theta + C}, \qquad (40)$$

where

$$a = \sum_j \|(\tilde{\chi}_n - \chi_{n-1})u_{j,n}\|_D^2,$$

$$b = \text{Re} \sum_j \langle \chi_{n-1} u_{j,n} - w_{j,n}, (\tilde{\chi}_n - \chi_{n-1})u_{j,n} \rangle_D,$$

$$c = \sum_j \|\chi_{n-1} u_{j,n} - w_{j,n}\|_D^2,$$

$$A = \sum_j \|(\tilde{\chi}_n - \chi_{n-1})u_j^{\text{inc}}\|_D^2,$$

$$B = \text{Re} \sum_j \langle \chi_{n-1} u_j^{\text{inc}}, (\tilde{\chi}_n - \chi_{n-1})u_j^{\text{inc}} \rangle_D,$$

$$C = \sum_j \|\chi_{n-1} u_j^{\text{inc}}\|_D^2, \qquad (41)$$

This is the quotient of two quadratics which, using elementary analysis, may be shown to attain its minimum when

$$\theta = \frac{-(aC - Ac)}{2(aB - Ab)}$$

$$+ \frac{\sqrt{(aC - Ac)^2 - 4(aB - Ab)(bC - Bc)}}{2(aB - Ab)}. \qquad (42)$$

This completes the description of the algorithm except for designating the starting values $w_{j,0}$. Observe that we cannot start with $w_{j,0} = 0$ since then $\chi_0^r = \chi_0^i = 0$ and the cost functional (29) is undefined for $n = 1$. Therefore we choose as starting values either the constant values that minimize the data error,

$$w_{j,0}^c = \frac{\langle f_j, G_j^S 1 \rangle_S}{\|G_j^S 1\|_S^2}, \qquad (43)$$

or the values obtained by backpropagation,

$$w_{j,0}^{\text{bp}} = \frac{\|G_j^{S^*} f_j\|_D^2}{\|G_j^S G_j^{S^*} f_j\|_S^2} G_j^{S^*} f_j. \qquad (44)$$

This completes the description of the algorithm.

6. COMPUTATIONAL TESTS

A number of tests have been done with the algorithm including stability tests, resolution tests and test for various kinds of contrast profiles [23]. Here some tests are presented for using the method for reconstructing AP-mines. We used one frequency of 500 MHz, a relative background permittivity of $\varepsilon_r = 5$ and a background conductivity of $\sigma = 0$ S/m. Further we used a computational domain of 29×29 cm. The AP-mines were given a relative permittivity of $\varepsilon_r = 7$ and a conductivity of $\sigma = 0$ S/m. Their diameter varied between 5 cm and 10 cm. The measurement curve, S, is a circle of radius 1 m and center at the center of D. The discrete form of the algorithm is obtained by dividing D into 29×29 subsquares, assuming the contrast, sources and fields are piecewise constant and the integrals over subsquares were approximated by integrals over circles of equal area which were calculated analytically [16]. The discrete spatial convolution of the operators G^D and G^{D^*} were computed using FFT routines [21]. The incident fields were chosen to be excited by line sources parallel to the axis of the scatterer. These sources were taken to be equally spaced on the measurement circle, and the source locations were also chosen as discretization points on the circle. All integrals on S were approximated by point collocation at the discretization points, that is, the rectangular rule with the integrand evaluated at the mid-points. The measured data were simulated by solving the direct scattering problem with a conjugate gradient method [21]. The circle S was subdivided into 30 equally spaced arcs. Each mid-point served as the location of a line source and all the mid-points served as receiver. In all test backpropagation has been used for the initial guess.

In figure 1, we show the original contrast profile of a circular mine with a 7 cm diameter.
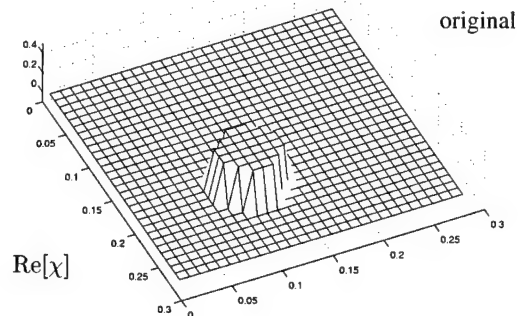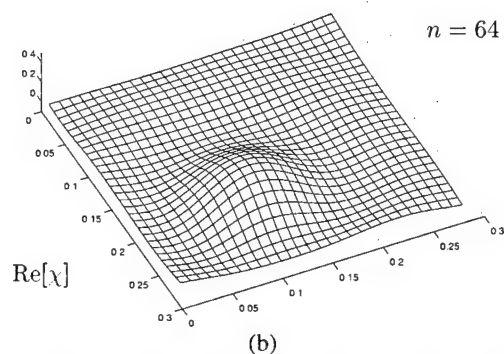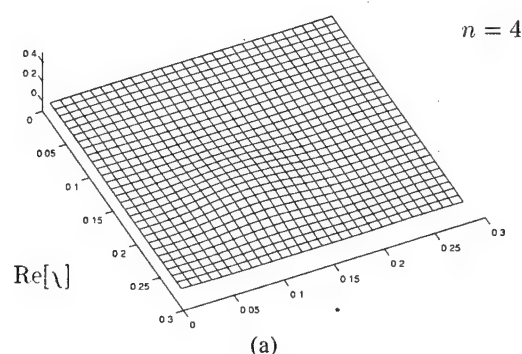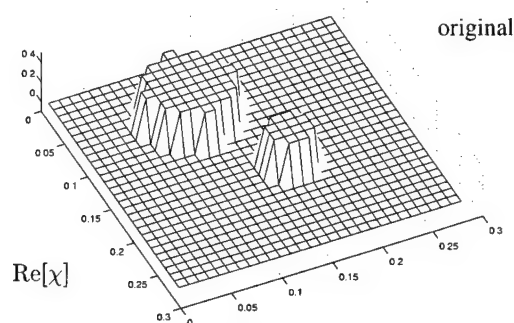

Fig 1: The original profile

(a)

(b)

Fig 2: The reconstructions after 4 iterations (a), and after 64 iterations (b), using CSI without *a priori* information.
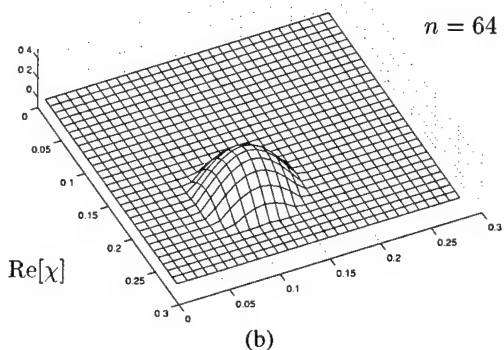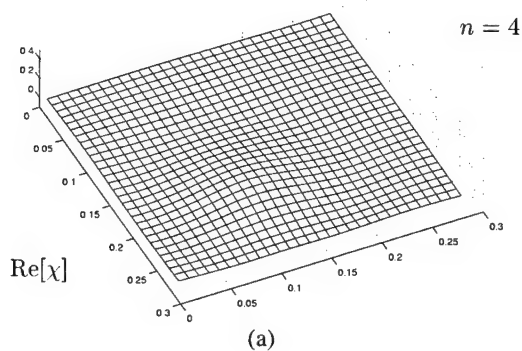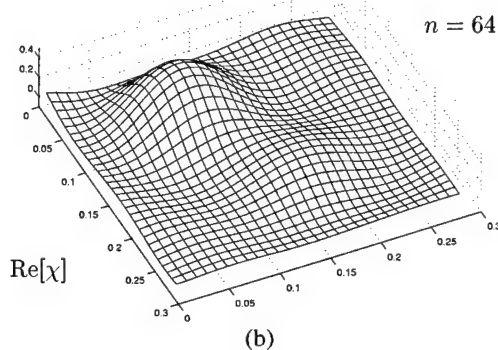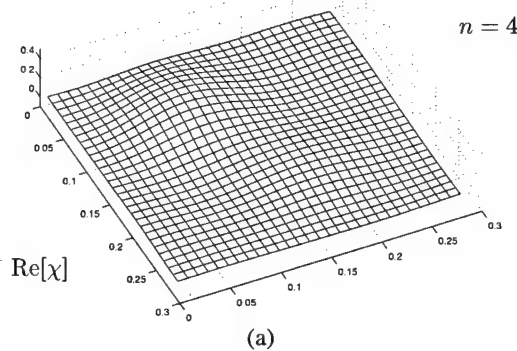

(a)

(b)

Fig 3: The reconstructions after 4 iterations (a), and after 64 iterations (b), using CSI with *a priori* information.

In figure 2, we show the reconstruction of the contrast profile of the mine after 4 and after 64 iterations. Continued iteration provided no noticeable improvement.

The conductivity of the mines is zero. We can therefore set the imaginary part of the contrast equal to zero. Since tests indicated that this restriction does not improve the algorithm, this restriction has been left out. Furthermore, since we used a permittivity of the mines which is higher than that of the background, we could use positivity for the mine permittivity. In figure 3, we show the reconstructed contrast profile after 4 iterations using CSI with *a priori* information, (a), and after 64 iterations with *a priori* information, (b). Each iteration took approximately 5 seconds on a Pentium PC computer.

Next we have put two mines in the configuration, one circular mine having a diameter of 5 cm and one having a

Fig 4: The original profile with two mines

(a)

(b)

Fig 5: The reconstructions after 4 iterations (a), and after 64 iterations (b), using CSI without *a priori* information.

(a)

(b)
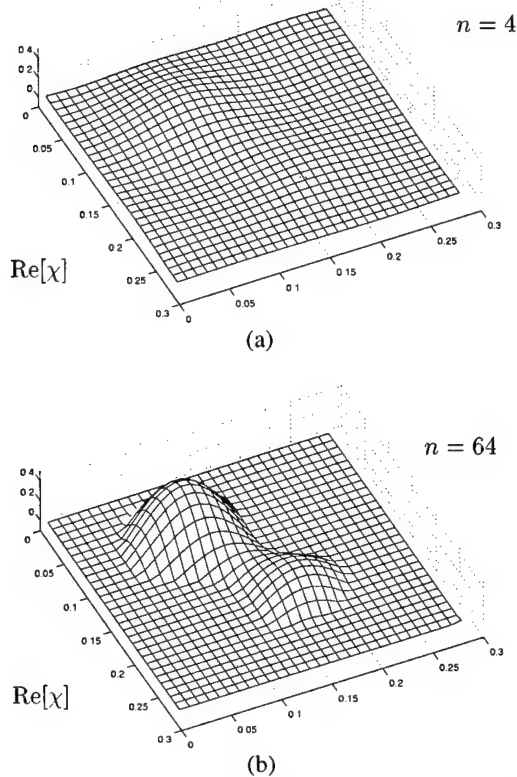
Fig 6: The reconstructions after 4 iterations (a), and after 64 iterations (b), using CSI with *a priori* information.

diameter of 10 cm. We have done the same tests. The results are in figures 4, 5 and 6. It is observed that the resolution in these figures is not as good as the resolution in the figures of the original paper. This is due to the size of the domain which is about $\frac{1}{2}\lambda \times \frac{1}{2}\lambda$. Higher frequencies, thus shorter wavelengths, will lead to an improved resolution but these waves have a lower penetration depth.

7. CONCLUSION

In this paper we have presented a new inversion algorithm (CSI) for profile reconstruction in acoustics and electromagnetics. The algorithm is based on the source type integral equation which relates measured data to a source distribution in the scattering object. The algorithm is akin to what Kohn and McKenney [15] call an alternating direction implicit (ADI) method wherein two sequences, one of sources and one of contrasts, are reconstructed iteratively by alternately updating the sources and the contrasts. Unlike the Kohn-McKenney method and other approaches based on the source type integral equation, e.g., Habashy *et al.* [11], the CSI algorithm does not involve completely solving the source type equation for each updated contrast. Similar to the modified gradient method, in each iteration there is no full inversion of the state equations involved.

A cost functional is defined consisting of errors in the source type equations and the state equations and the updates in the sources are found as a conjugate gradient step after which the contrast is updated by minimizing the error in the state equations which can be done very simply. The source updates are similar in spirit to those used in the modified gradient method while the contrast updates are found in a simple fashion in which *a priori* information is easily included. A number of numerical tests indicate that

this new algorithm exhibits the best features of the modified gradient algorithm, successfully reconstructing a variety of contrasts and fairly insensitive to noise. However, the new algorithm exhibits additional properties which surpass the modified gradient approach. It is faster, requires less memory as well as less data and more easily accommodates *a priori* information.

To give some idea of the computational complexity, if J denotes the number of excitations and N denotes the number of subdomains in the test domain, then the time required for each iteration is roughly $2J$ times the time for one step in the conjugate gradient solution of the forward problem for one excitation, with a memory requirement of approximately $5JN \times 16$ bytes (complex double precision). The time required for each iteration is roughly one third that needed in the modified gradient algorithm with no *a priori* information on the contrast and is an order of magnitude faster if positivity is included for both real and imaginary parts.

No tests have yet been carried out on the effect of additional regularizers such as total variation which proved effective for the modified gradient algorithm [24]. This is one of the subjects for future work. The simplicity, speed and reduced memory requirements offer hope that this technique will provide a feasible approach to three-dimensional inversion problems.

REFERENCES

1. Bleistein, N., and Cohen, J.K., "Nonuniqueness in the inverse source problem in acoustics and electromagnetics", J. Math. Phys. 18, 1977, pp 194–201.

2. Bojarski, N.N., "Inverse scattering inverse source theory", J. Math. Phys. 122, 1981, pp 1647-50.

3. Bojarski, N.N. Comments on "Nonuniqueness in inverse source and scattering problems", IEEE Trans. Antennas Propagat. 30, 1982, pp 1037–38.

4. Caorsi S., Gragnani, G.L. and Pastorino, M., "A multiview microwave imaging system for two-dimensional penetrable objects", IEEE Trans. Microwave Theory Tech. 38, 1991, pp 845-51.

5. Caorsi S., Gragnani, G.L. and Pastorino, M., "Numerical solution to three-dimensional inverse scattering for dielectric reconstruction purposes", IEE Proc. H 139, 1992, pp 45-52.

6. Chew, W.C. and Wang, Y.M., "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method", IEEE Trans. Med. Imag. 9, 1990, pp 218–25.

7. Chew, W.C., Wang, Y.H., Otto, G., Lesselier, D. and Bolomey, J.Ch., "On the inverse source method of solving inverse scattering problems", Inverse Problems 10, 1994, pp 547-53.

8. Devaney, A.J., "Nonuniqueness in the inverse scattering problem", J. Math. Phys. 19, 1978, pp 1526–31.

9. Devaney, A.J. and Sherman, G.C., "Nonuniqueness in inverse source and scattering problems", IEEE Trans. Antennas Propagat. 30, 1982, pp 1034–37.

10. Habashy, T.M., Chow, E.Y. and Dudley, D.G., "Profile inversion using the renormalized source-type in-

tegral equation approach", IEEE Trans. Antennas Propagat. 38, 1990, pp 668–81.

11. Habashy, T.M., Oristaglio, M.L. and De Hoop, A.T., "Simultaneous nonlinear reconstruction of two-dimensional permittivity and conductivity", Radio Sci. 29, 1994, pp 1101–18.

12. Kleinman, R.E. and Van den Berg, P.M., "A modified gradient method for two-dimensional problems in tomography", J. Computat. Appl. Math. 42, 1992, pp 17–35.

13. Kleinman, R.E. and Van den Berg, P.M., "An extended range modified gradient technique for profile inversion', Radio Sci. 28, 1993, pp 877–84.

14. Kleinman, R.E. and Van den Berg, P.M., "Two-dimensional location and shape reconstruction", Radio Sci. 29, 1994, pp 1157–69.

15. Kohn, R.V. and McKenney, A., "Numerical implementation of a variational method for electrical impedance tomography", Inverse Problems 6, 1990, pp 389–414.

16. Richmond, J.H., "Scattering by a dielectric cylinder of arbitrary cross section shape", IEEE Trans. Antennas Propagat. 13, 1965, pp 334-41.

17. Roger, A., "A Newton-Kantorovich algorithm applied to an electromagnetic inverse problem", IEEE Trans. Antennas Propagat. 29, 1981, pp 232–38.

18. Stone, W., "Comments on "Nonuniqueness in inverse source and scattering problems", IEEE Trans. Antennas Propagat. 30, 1982, pp 1037–38.

19. Tabbara, W., Duchêne, B., Pichot, Ch., Lesselier, D., Chommeloux, L. and Joachimowicz, N., "Diffraction tomography: Contribution to the analysis of applications in microwaves and ultrasonics", Inverse Problems 4, 1988, pp 305–31.

20. Tijhuis, A.G., "Born-type reconstruction of material parameters of an inhomogeneous lossy dielectric slab from reflected-field data", Wave Motion 11, 1989, pp 151–173.

21. Van den Berg, P.M., "Iterative computational techniques in scattering based upon the integrated square error criterion", IEEE Trans. Antennas Propagat. 32, 1981, pp 1063–71.

22. Van den Berg, P.M. and Haak, K.F.I., "Profile inversion by error reduction in the source type integral equations", Wavefields and Reciprocity, edited by P.M. van den Berg, H. Blok and J.T. Fokkema, Delft University Press ISBN 90-407-1402-9, 1996, pp 87–98.

23. Van den Berg, P.M., and R.E. Kleinman, "A contrast source inversion method," Inverse Problems 13, 1997, pp. 1607-1620.

24. Van den Berg, P.M. and Kleinman, R.E., "A total variation enhanced modified gradient algorithm for profile reconstruction", Inverse Problems 11, 1995, pp L5–10.

25. Van den Berg, P.M., Coté, G. and Kleinman, R.E., ""Blind" shape reconstruction from experimental data", IEEE Trans. Antennas Propagat. 43, 1995, pp 1389–96.

Mine Detection with Microwaves

M. Magg

IABG mbH, Einsteinstr. 20, D-85521 Ottobrunn, GE

J. Nitsch

Otto-von-Guericke-University Magdeburg, P.O. Box 4120, D-39016 Magdeburg, GE

Summary

Location and identification of buried land mines is a real challenge for sensor technology and target identification algorithms. We analyse the performance of a bistatic microwave imaging system with a focused synthetic aperture. If the soil is homogeneous and dry with a very smooth surface it will be possible to identify even plastic mines under a 10 cm overburden. However, under a rough surface or in wet soil even a relatively big metallic anti-tank mine could be missed, since the signal to clutter ratio gets quite poor under these circumstances.

1. Introduction

Military conflicts left vast tracts of country under the suspicion of contamination by land mines. Efficient, reliable mine detecting, identification and clearing techniques are urgently needed. Buried land mines present a particular hard problem. Standard detection procedures for buried land mines still rely on metal detectors and bayonet probing. Brute force clearing uses deep plowing, moulding and harrowing. Obviously, these procedures can be applied only on limited areas.

The buried mines range from small, inexpensive anti-personal plastic mines with very little metal content to bulky metallic anti-tank mines. Possible mine shapes include disks, cylinders, square boxes and more fancy geometries. The sizes of mines start at about 5 cm and may go up to 40 cm. Looking from above ground, it is virtually impossible to find a unique signature that can be used to discriminate a buried mine from other similar looking pieces of scrap. A low false alarm rate, however, is required for an efficient clearing of expanded lots suspected of mines.

The requirements on civilian mine clearing are quite ambitious:

- Detection rate of 99.9% or better,
- Operating at rough soil surface as well as for wet soil,
- Low false alarm rate, i.e. clear distinction of mines versus scrap, rock pieces, soil surface structures.

It was suggested to exploit the explosive charge which is contained in every mine to detect them under a shallow soil overburden. Almost all explosive charges consist of strongly nitrogenious chemicals. Therefore a high local nitrogen concentration in the soil may indicate a buried mine. The proposals for spotting local nitrogen concentrations close to the soil surface require variing technical expenditure. Sniffing by specially trained dogs and γ-ray activation of the natural nitrogen isotop with subsequent detection of the resulting instable isotop are two examples that have been suggested [1].

Another starting point for the detection and identification of buried mines hopes to exploit the casing geometry of buried mines as seen by microwaves. An aspect independent way of perceiving the geometry of a target is given by the complex frequency poles of the electro-magnetic scattering amplitude caused by a buried target if it gets illuminated by microwaves [2]. If one has a catalogue of these complex frequency pole patterns for all interesting buried targets one may hope to determine the target from an analysis of microwave scattering data in the relevant frequency band.

One diffculty with this method when applied to targets buried in the soil is the variing influence of the soil overburden on

the received scattering signal. This is in contrast to the better known application of this identification technique to airborne targets. The influence of the air on the back scattering data is much smaller and reasonably well under control. The effect of soil on the comlex resonance frequencies of a buried target varies with the thickness and composition of the soil . overburden. Furthermore in the S-band, which is most relevant for mine identification, wet soil is expected to produce a strong signal dispersion, which may disturb the final complex pole analysis of the scattering data.

In this paper we analyse narrow band microwave imaging of buried objects as inspired by much employed high resolution radar techniques [3,4]. The objective is to investigate the performance of microwave imaging for mine searching and identification, exhausting the physical capabilties of microwaves without caring much about implementation details.

2. Bistatic Microwave Imaging with Focussing Synthetic Aperture

In order to obtain a good spatial resolution for buried targets the wave length used for microwave imaging should be chosen as small as possible. Since a mine buried under a 20 cm thick layer of moderately wet soil should be still visible, the microwave should be able to cross a 1 cm layer of water without too much absorption. Looking at the absorption coefficient of water as a function of frequency one learns, that one cannot go beyond S-band frequencies. Therefore spatial resolution for ground penetrating imaging systems based on non-ionizing electromagnetic waves cannot go much below 4 to 6 cm. All numerical studies in this paper use a frequency of 3.5 GHz.

In order to look into the soil one has to overcome the reflection from the soil surface which is superimposed to the scattering echos from buried objects. For vertical illumination a soil surface produces a reflection factor r , given by:

$$r = (n-1)/(n+1),\qquad(1)$$

where n is the index of refraction of the soil. This means for $n = 3$ half of the incident wave amplitude would be reflected at the ground surface and would interfere with the echo from buried objects. Since the radar range resolution is not sufficient to separate the ground surface echo from that of shallowly buried targets, suppression of the ground surface echo is required.

The reflection from a plane surface is minimal (ideally zero) if the plane is illuminated along the Brewster direction by a wave which is polarized in the plane of incidence. The Brewster angle is determined by the index of refraction of the soil:

$$\tan\theta_B = n\qquad(2)$$

For soil this angle lays in the range from 52° to 85°. If the illumination misses the Brewster angle by 3° the reflection factor will be still less than 10%.
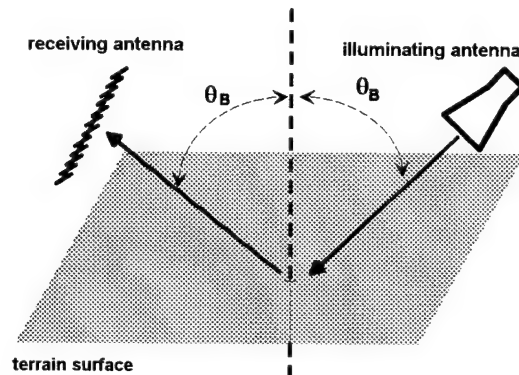
Fig. 1: Bistatic microwave mine detection and imaging system

The spatial resolution increases with the aperture dimension of the receiving antenna and decreases with the distance from the antenna to the focussed region. Hence, in order to achieve a high resolution power for the microwave imaging system the receiving aperture should be as close to ground as possible. Practical operation conditions require a sufficient terrain clearance. In our study we assume that the receiving antenna is 0.7 m above average

ground level. The receiving antenna is designed as an approximately 1 m long linear dipole array. A much bigger array would restrict the possible use of the mine searching system which could be mounted in front of a small cross-country vehicle.

The resolution in the direction orthogonal to the linear array is produced by the forward move of the complete imaging system. The receiving array sweeps a two dimensional ground strip, thereby forming a synthetic aperture, much like in a forward looking SAR-system. At every step, the received complex scattering amplitude as seen by the individual dipole elements of the receiving array will be stored in the memory of a data processing unit.

The stepping direction will be called 'longitudinal' whereas the orthogonal direction, parallel to the linear receiving antenna, will be addressed as 'lateral'.

The longitudinal length of the synthetic aperture, i.e. the number of real antenna positions used to compose an image of a point in real space determines the resolution of the system in the longitudinal direction. The physical length of the array, i.e. the lateral width of the synthetic apertur controls the lateral resolution.

Since the microwave scattering is concentrated around the reflection angle, an aperture length in stepping direction of the order of 1 m would be sufficient. Larger aperture dimension would improve only marginally the resolution power of the system at the expense of a longer image processing time.

Figure 1 illustrates the basic system design. Illuminating and receiving antenna are kept at a riggid horizontal distance from each other. In our study this distance is 2 m. The complete system steps forward in the direction from the receiving array to the illuminating horn. The receiving array is oriented laterally. The horn and the dipole elements of the receiving array both are inclined by angle θ_B against the vertical.
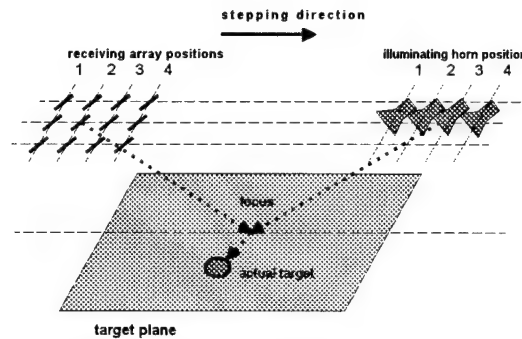
Fig.2: Synthetic aperture with 4 real antenna positions.

Figure 2 sketches the bistatic system in action, forming a synthetic plane aperture which consists out of 4 real antenna positions.

3. Longitudinal and Lateral Resolution Power

The processing of the stored scattering data consists in focussing the data relative to a grid of <u>focus points</u> which is projected onto the presumed mine laying plane (target plane). For a chosen focus point in the target plane one calculates the optical path length from the illuminating antenna via that focus point to the individual antenna elements in the receiving aperture. The complex scattering amplitudes as actuallly received by the elements of the synthetic aperture get multiplied by the complex conjugate phases related to the path length from illuminating antenna to receiving element via the chosen focal point on the target plane. The resulting phase corrected amplitudes are summed over all elements of the synthetic aperture. For a target that sits directly at a focus point all phase-corrected amplitudes of the individual elements in the synthetic aperture add up whereas the phase-corrected elementwise scattering amplitudes of a target which is more than a resolution distance away from that focus point would cancel each other almost completely in the summed antenna response.

This focussing is performed one by one for all focus points of the grid. By mapping the distribution of the focussed antenna

response onto the target plane one obtains an image of all scattering objects which are buried within a certain layer around the target plane. The thickness of this layer is determined formally by the depth of fields of the imaging system. As we shall see later it is not the formal depth field but the much shorter electromagnetic penetration depth which puts actually the limit for the vision into the soil.

Mathematically the focussing algorithm is defined as follows. Let $A(m,n)$ be the complex scattering amplitude as actually received by the m-th array element of the array at the n-th step position over the target plane. Furthermore let (x,y) be the coordinates of a chosen focus point on the target plane. Then the total amplitude received by the focussed aperture is given by:

$$A_f(x,y) =$$
$$\sum_{m,n} A(m,n) \cdot exp\{-j \cdot k$$
$$[\sqrt{h^2 + (x_n - x)^2 + (y_m - y)^2}$$
$$+ \sqrt{h^2 + (x_n + 2a - x)^2}]\}$$

(3)

h   distance of the target plane from the aperture plane,

k   wave vector $2\pi \cdot f / c$ (f = frequency, c = speeed of light)

a   half distance between illuminatinig horn and receiving linear array,

x_n longitudinal position of the receiving antenna at the n-th step,

y_m lateral distance of the m-th array element from array line center.

The height h and the distance a are related by the incident angle:

$$\cot \theta_B = h / a \quad . \quad (4)$$

The focussing is achieved by compensating the phases as actually received by the different antenna elements by the path lengths of the ray pencil which passes the chosen focus point.

The summation in (3) runs over all elements m of the real linear array and all positions n which form the synthetic aperture. Note that the focus point belongs

to the Fresnel region of the receiving antenna. The distance from the receiving antenna to the focus point is not much larger than the longitudinal and lateral aperture dimension.

The formula (3) neglects the fact, that the focus point is below ground. The correct electromagnetic path from transmitting horn to receiving array antenna via focus point would have to take refraction into account. For a shallow overburden this would cause a small path length correction. In the numerical simulation, which we report later on, the full 'optical' path length has been computed and used for the focussing. Unfortunately the explicit expression for the full optical path length is somewhat bulky to write down explicitely, therefore we stay with the deputy formula (3).

One may estimate the spatial resolution of the synthetic aperture following a popular $\lambda/2$-argument in optics and antenna theory:

First compute the path length for the so-called central ray. The central ray starts at the wave source, scatters off the target point and ends at the centre of the receiving aperture. Next compute the path length for the so-called boundary ray. The boundary ray shares with the central ray the same starting and scattering points but it ends at the boundary of the receiving aperture. If both path lengths differ by half a wavelength, then the scattering signal caused by that target point cancels in the received total antenna signal.

For a focussed array antenna it is the relative path length which enters the preceeding argument. The relative path length is obtained from the original path length by subtracting from it the length of the focus reference path. The focus reference path has the same starting and end points as the original central ray, resp. boundary ray, however, it passes through the focus point instead of being scattered off at the target point.

According to the $\lambda/2$ argument a target point close to a chosen focus point will not contribute significantly to the signal

received from that focus point, if the relative path lengths for the central and the boundary ray via that target point differ by half a wavelength or more.

Figure 3 displays the rays which are relevant for the estimation whether a point T contributes to the image of an object at the focus point F or not.
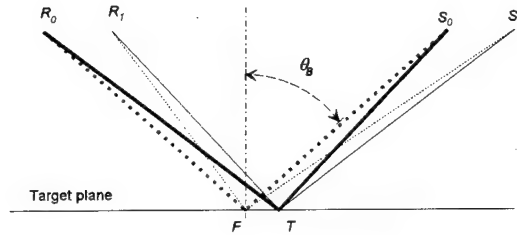

Fig.3:   Cut in longitudinal direction: central and boundary rays of the synthetic aperture; R_0 and R_1 central and boundary receiver antenna position, respectively, S_0 and S_1 source central and boundary source position, respectively, T target position , F focus position

Here and in the figures below we distinguish the rays types by the convention:

central rays   → fat lines,
boundary rays → normal lines,
rays scattered → solid lines,
at the target
reference rays → broken lines.
passing the
focus

The distances $\overline{R_0 R_1}$ and $\overline{S_0 S_1}$ are both equal to half of the longitudinal dimension L of the synthetic aperture. Let x denote the distance from the target T to the chosen focus F. The horizontal distance between the receiver array and illuminating horn is $2a$. Following Figure 3 one identifies $\overline{R_0 T} + \overline{T S_0}$ as the path length for the central ray scattered at the target, similarly $\overline{R_0 F} + \overline{F S_0}$ is the path length for the associated reference ray. The length of the boundary ray scattered at the target is

$\overline{R_1 T} + \overline{T S_1}$, and finally $\overline{R_1 F} + \overline{F S_1}$ is the path length for the reference ray associated with the boundary ray.

The path length difference between the signals received by the boundary element and the central element both focussed relative to point F is composed as follows:

$$W(x) = \overline{R_0 T} + \overline{T S_0} - \overline{R_0 F} - \overline{F S_0}$$
$$- \overline{R_1 T} - \overline{T S_1} + \overline{R_1 F} + \overline{F S_1}$$

In linear approximation for x this path length difference becomes:

$$W(x) \approx \frac{x \cdot L}{\sqrt{h^2 + a^2}}$$
$$\cdot \left[\frac{h^2}{a^2 + h^2} + \frac{L^2}{8} \cdot \frac{5a^2 - h^2}{(a^2 + h^2)^2} \right] \qquad (5)$$

The factor which multiplies the distance x in equation (5) corresponds to the 'numerical aperture' in optics. For a rough estimate one may neclect non-linear corrections in the longitudinal aperture dimension L. The contribution of the target point to the signal as received by the synthetic aperture focussed at point F vanishes if the path length difference W is equal to half a wave length λ. This gives an estimate of the smallest longitudinal scale x_{min} which can be resolved by the the bistatic synthetic focused aperture system:

$$x_{min} \approx \tfrac{1}{2} \cdot \lambda \cdot \sqrt{h^2 + a^2} / (L \cdot \cos^2 \theta_B) \qquad (6)$$

Except for an additional $\cos \theta_B$ factor in the denominator, this is just what is known from synthetic apertures used in radar applications. Note the factor 1/2 which is a bonus not valid for real apertures.
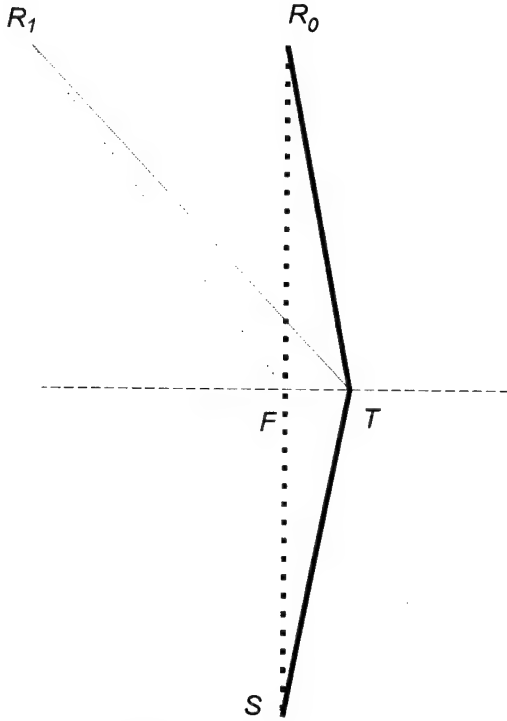
Fig. 4: Lateral direction: central and boundary rays of the bistatic imaging system; R_0 and R_1 central and boundary receiver array element, respectively, S position of the illuminating source.

In the lateral direction the synthetic aperture is actually a real aperture given by the length of the receiving array line. Figure 4 sketches the ray geometry as seen from above.

The linear receiving array antenna has a length $2\,B$. The relevant distance scales in Figure 4 are:

$$\overline{R_1\,R_0} = \tfrac{1}{2}B;$$

$$\overline{R_0\,F} = \overline{F\,S} = \sqrt{a^2 + h^2}\;;\quad \overline{F\,T} = y \qquad (7)$$

Remember that R_1 and R_0 keep a distance h to the target plane. The focus point F and the target point T lay in that target plane. The path length difference between the signal received by the lateral boundary element and the central element both focussed relative to point F one obtains:

$$W(y) = \overline{R_0 T} + \overline{TS} - \overline{R_0 F}$$
$$- \overline{FS} - \overline{R_1 T} - \overline{TS} + \overline{R_1 F} + \overline{FS}$$

The source position cancels exactly in the path length difference. Up to non-linear terms in the lateral target distance from the focus point one gets:

$$W(y) \approx y \cdot B \,/\, (2\sqrt{a^2 + h^2 + \tfrac{1}{4}B^2}\,) \qquad (8)$$

The analogous reasoning as for the longitudinal case implies a lateral resolution:

$$y_{\min} \approx \lambda \cdot \sqrt{a^2 + h^2}\,/\,B \qquad (9)$$

Non-linear contributions of the lateral apertur dimension B have been ignored in this estimate.
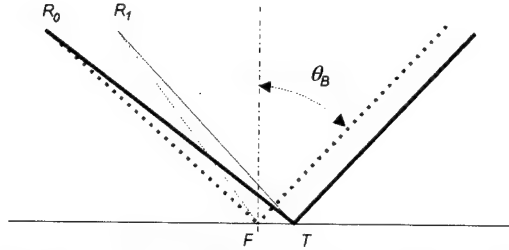
Fig. 5: Longitudinal direction: central and boundary rays in the synthetic aperture; the illumination is by a remote spatially fixed source. R_0 receiver antenna central position, R_1 receiver antenna boundary position, T Target point , F focus point

The lateral resolution changes if the illumination occurs by a spatially fixed transmitting antenna, i.e., if the horn antenna does not follow the forward stepping of the receiving array. Figure 5 shows the situation if the ground gets illuminated by a remote spatially fixed transmitting antenna.

In this environment the path length difference between the signal received by the lateral boundary element and the central element both focussed relative to point F is:

$$W(x) \approx \frac{1}{2} \cdot \frac{x \cdot L}{\sqrt{h^2 + a^2}} \cdot \frac{h^2}{a^2 + h^2}$$
$$\cdot \left[1 - \frac{3}{4} \cdot \frac{a \cdot L}{a^2 + h^2} \right] \qquad (10)$$

This leads to the following estimate for the longitudinal resolution:

$$x_{\min} \approx \lambda \cdot \sqrt{a^2 + h^2}\,/\,(L \cdot \cos^2 \theta_B) \qquad (11)$$

This is twice as coarse as what is obtained with the comoving illuminating antenna.

Finally the depth of fields for the bistatic imaging system has to be determined. If the receiving antenna focusses at objects in a chosen target plane (parallel to the ground

surface) it displays also objects which lay somewhat above or below that target plane. The depth range for which targets are reasonably well mapped by the bistatic imaging system (while focussing at a chosen target depth) is the 'depth of fields'.
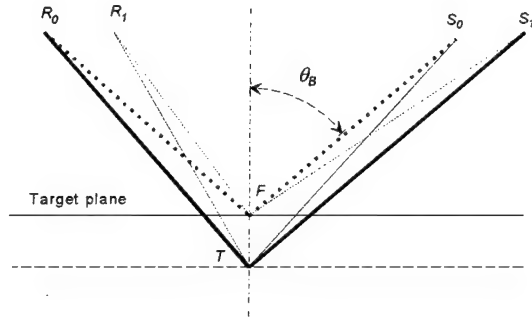

Fig.6: Depth of fields: central and boundary rays for a synthetic aperture; actual target T is below focus point F.

Figure 6 shows the central rays and the boundary rays for a target point laying in a different depth than the chosen focus.

If the antenna is focussed at point F and the scattering occurs at target T, then the path length difference becomes:

$$W(z) = \overline{R_1 T} + \overline{TS_1} - \overline{R_1 F} - \overline{FS_1}$$
$$- \overline{R_0 T} - \overline{TS_0} + \overline{R_0 F} + \overline{FS_0}$$

The distance $\overline{F\ T}$ of the actual target from the assumed target plane is called z. Keeping only linear terms in z one finds:

$$W(z) = h \cdot z \cdot [1 / \sqrt{(a + \tfrac{1}{2}L)^2 + h^2}$$
$$+ 1 / \sqrt{(a - \tfrac{1}{2}L)^2 + h^2} - 2 / \sqrt{a^2 + h^2}\ ] \quad (12)$$

Neglecting quartic and higher terms in the longitudinal aperture dimension L one gets:

$$W(z) \approx \frac{z}{4} \cdot \frac{L^2}{a^2 + h^2} \cdot \cos\theta_B \cdot [2 - 3\cos^2 \theta_B] \quad (13)$$

The contribution of an actual target to the assumed target plane can be ignored, if the distance of the actual target to that plane is larger than:

$$z_{min} \approx 2\lambda \cdot (h^2 + a^2) / [L^2 \cdot \cos\theta_B \cdot (2 - 3\cos^2\theta_B)] \quad (14)$$

This results expresses the limits of the depth of field which are due to the longitudinal aperture dimension L. The lateral aperture dimension B causes a slightly different bound for the depth of field:

$$z_{min} \approx 4\lambda \cdot (h^2 + a^2) / (B^2 \cdot \cos\theta_B) \quad (15)$$

Actually the formal depth of field is not particularly relevant for a ground penetrating microwave system. In general the penetration depth of 3.5 GHz microwaves in soil is much shorter ($\approx 3\lambda$) than the formal depth of field which was estimated by eqs. (13) and (14).

Figure 7 shows the calculated intensity distribution caused by a single pointlike target as seen by a bistatic imaging sytem focussed at the target plane. The intensity is given by the modulus squared of the amplitude in eq. (3). The pointlike target is modelled as an isotropical scattering center in the target plane at coordinates (x', y') which gets illuminated by a comoving transmitting antenna at a horizontal distance $2a$ from the receiving array line position x_n:

$$A(m,n) = exp\{j \cdot k$$
$$\cdot [\sqrt{h^2 + (x_n - x')^2 + (y_m - y')^2} \quad (16)$$
$$+ \sqrt{h^2 + (x_n + 2a - x')^2}\ ]\}$$

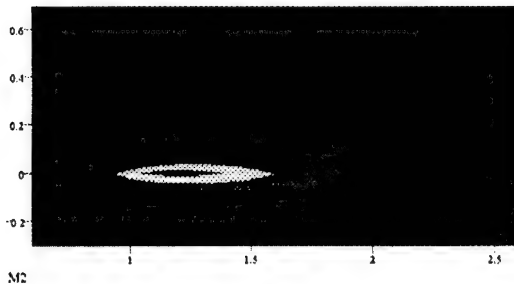No approximations have been made in the evaluations of the square roots.
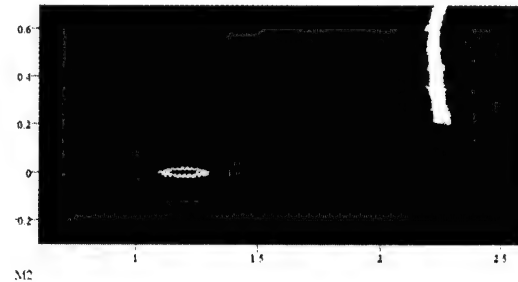
Fig. 7: Image of a pointlike target at position (1.2,0) in the target plane, 0.75 m below the aperture plane of a bistatic synthetic aperture imaging system, illuminating antenna moves together with the receiving array.

The longitudinal resolution shown is about 0.3 m, whereas the lateral resolution is 0.06 m. According to the eqs. (6) and (9), an aperture size of 40 cm by 120 cm allows a resolution of 48 cm by 9 cm (longitudinal by lateral), at a target plane 0.75 m below the receiving array.

Consider now a bistatic system where the illuminating horn remains fixed while the receiving array is sweeping the synthetic aperture. If the illuminating horn is not too close to the surface, the illumination can be modelled as a portion of a plane wave incident on the surface under the angle θ_B. One obtains an expression similar to equation (3) for the total amplitude received by the focussed aperture of that system. The elementwise received amplitudes now take the form:

$$A(m,n) = exp\{j \cdot k$$
$$\cdot [\sqrt{h^2 + (x_n - x')^2 + (y_m - y')^2} \quad (17)$$
$$- x' \cdot sin\theta_B]\}$$

The focussing phase shift is

$$exp\{-j \cdot k \cdot [\sqrt{h^2 + (x_n - x')^2 + (y_m - y')^2}$$
$$- x' \cdot sin\theta_B]\} \quad (18)$$

The intensity received by this system design from the same pointlike target is now displayed in Figure 8. Note that the longitudinal resolution is now at least twice as coarse as for the system with comoving illumination. This behaviour is predicted by eq. (11) as compared to eq. (6).

Fig. 8: Image of a pointlike target at position (1.2,0) in the target plane 0.75 m below the aperture plane of a bistatic synthetic aperture imaging system; the illuminating antenna is spatially fixed

The analysis of resolution power was based solely on the array pattern characteristics. The directional pattern of the individual elements in the receiving antenna array was not taken into account. Furthermore the isotropic scatterer model of the buried target is a severe over-simplification. The bistatic scattering cross-section of a buried target tends to concentrate the scattering field around the reflection direction. Both directional gains would reduce the aperture dimensions which are required in order to obtain a desired resolution.

4. Signal to Clutter Ratio

As important as the resolution power is the targets contrast relative to the background for the imaging device. Since an ideally flat ground surface would cause no reflection the background is produced solely by surface structures and inhomogeneities in the soil. To get a feeling for realistic signal to clutter ratios when the bistatic imaging system is applied to mine searching, we display first the signals of mine-like objects buried under an ideal overburden, e.g. homogeneous sand with a smoothly raked surface. Next we compare this ideal case with the image produced by a distinctively structured soil surface without any buried mines.

4.1 Scattered Fields of Buried Mines and of Soil Surface Structures

In this paragraph we display the scattering fields which are caused by various buried targets and soil surfaces when the surface gets illuminated by an incident plane wave. The incident angle for the illumination is always $\theta_B = 58°$. The incident field is normalized to 1 V/m, the frequency is 3.5 GHz.

The scattering fields are obtained by employing numerical simulation method as provided by a well established finite differences time domain code [5].

Two types of soil are investigated:

- very dry sand with dielectric constant $\varepsilon'=2.55$ and specific conductivity $\kappa=0.001\ S/m$,
- wet clay with dielectric constant $\varepsilon'=10$ and specific conductivity $\kappa=0.1\ S/m$.

For both soil types inhomogeneities are not included in the model. However, the model for the plastic mine also may be taken as a hint of the kind of echo that is to be expected from a natural inhomogeneity in the soil. The mine models are buried under a 10 cm deep uniform soil layer. The scattered fields are plotted in the plane of incidence and in a horizontal cut 0.65 m above average ground level. The horizontal plot displays only one half of the scattering field which is symmetric relative to the incident plane, since all targets in this study have this symmetry and the polarisation is parallel to the incident plane.

The contour plots (Figs. 9 and 10 to 15) linearly map the intensity of the scattered field into gray values. White corresponds to the maximum intensity, black to zero intensity. The scales at the plot axis state the longitudinal, the vertical and the lateral distances in meter. The mines are indicated in the incident plane plots, close to longitudinal coordinate 0. The plots display the scattered fields only, the incident field coming from top right has been filtered out. The concentration of the scattering in the reflection direction is clearly visible. Due to available computer resources the modelled soil layer was limited to 0.5 m in depth. For the dry sand this is somewhat less than what is desirable. Therefore the displayed field distribution in the lower part of the dry sand layer has to be taken with some care. The scattered field in the airspace, however, is believed to be simulated reliably.

metallic anti-tank mine in dry sand	0.23
big plastic mine in dry sand	0.065
rough terrain surface in dry sand	0.35
square metallic mine in dry sand	0.055
metallic anti-tank mine in wet clay	0.038
rough terrain surface in wet clay	0.35

Tab. 1: Peak values of scattered fields at 0.65 m above ground level

Table 1 summarizes the peak values of the microwave scattering signals as seen by a probe 0.65 m above average ground level for the scattering objects which have been examined in this study.

The modelling data of the employed mine types are as follows:

metallic anti-tank mine: cylinder, $\varnothing$ 24 cm, height 8 cm, ideally conducting, big dielectric mine: cylinder, $\varnothing$ 24 cm, height 8 cm, non-conducting, $\varepsilon'= 3.7$, metallic square mine: 10 cm squared, height 4 cm, ideally conducting.

The modelling of the rough surface is represented in Figure 11. The displayed surface segment is 3 m long.
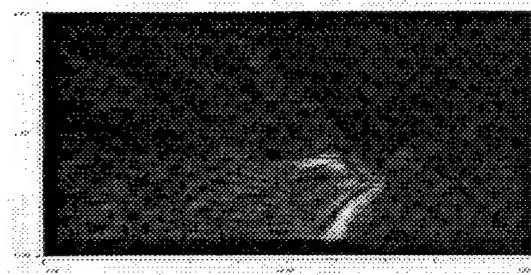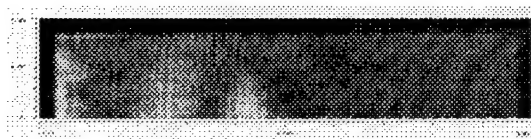

(a) max=1.2 V/m

(b) max=0.26 V/m
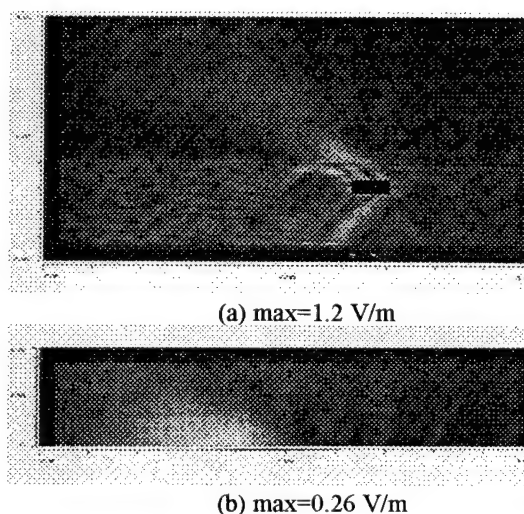
Fig. 9: Scattered field produced by a metallic cylindrical mine, $\varnothing$ 24 cm, height 8 cm buried in dry sand; (a) incidence plane, (b) horizontal plane.
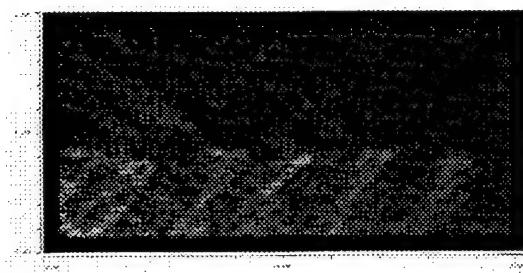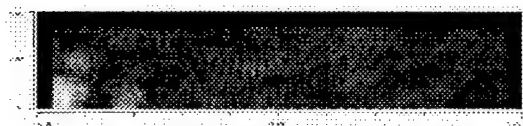
(a) max=1.8 V/m

(b) max=0.072 V/m

Fig. 10: Scattered field produced by a dielectric cylindrical mine, $\varnothing$ 24 cm, height 8 cm, ε'=3.70, buried in dry sand; (a) incidence plane, (b) horizontal plane

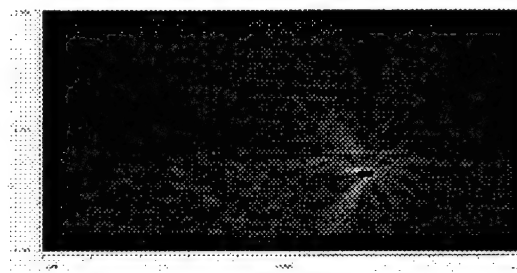Fig. 11: Coputer modell of a rough terrain surface with sharp relief structures ±5 cm.
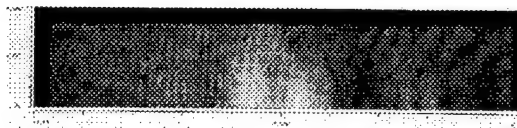

(a) max=2.8 V/m

(b) max=0.6 V/m

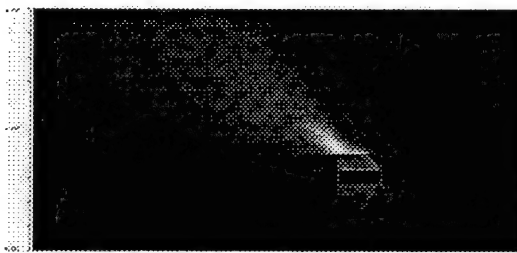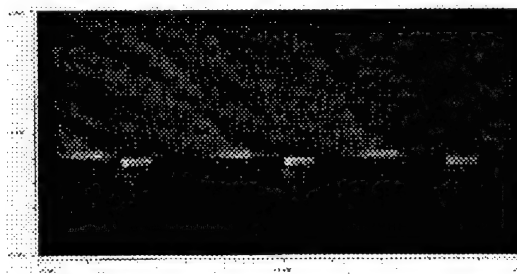Fig. 12: Scattered field produced by a rough terrain surface in dry sand; (a) incidence plane, (b) horizontal plane

(a) max=0.76 V/m

(b) max=0.06 V/m

Fig. 13: Scattering field produced by metallic 10 cm square mine, height 4 cm buried in dry sand; (a) incidence plane, (b) horizontal plane

(a) max=0.18 V/m

(b) max=0.04 V/m

Fig. 14: Scattered field produced by a metallic cylindrical mine, $\varnothing$ 24 cm, height 8 cm buried in wet clay; (a) incidence plane, (b) horizontal plane

(a) max=1.4 V/m

(b) max=0.45 V/m

Fig. 15: Scattering field produced by a rough terrain surface in wet clay; (a) incidence plane, (b) horizontal plane

4.2 Microwave Images of Buried Mines and of Soil Surface Structures

The scattered field as displayed in the previous section is the input data for the computational image formation as outlined in Section 3. Note that in the numerical simulation the illumination was mounted at a fixed position in space during the recording of the scattered field. The reason for this set-up was purely technical: It is much cheaper (i.e. faster) to simulate a spatially fixed illumination, since it requires the computation of just one scattered field configuration. The original bistatic, synthetic aperture imaging system uses a comoving illuminating horn antenna. This implies for every single step forward a separate computation of the scattered field, even though this field is evaluated only at the corresponding position of the receiving array elements.

For the signal to clutter ratio the two different methods are expected to give the same results. As to the resolution power it was argued in Section 3 , that the spatially fixed illumination needs a longitudinal aperture roughly twice as large as the comoving illumination in order to produce the same longitudinal resolution. The images in this section are obtained with the following aperture dimensions: 0.8 m lateral, 0.32 m longitudinal.

To evaluate the following images of targets as produced by the studied bistatic imaging system-design one has to refer to the gray values scales given on the left side of the plots. The contrast produced by the metallic anti-tank mine is almost 6-times stronger than that of the big plastic mine. Comparing the gray values of the mine models with those of the surface models one may predict a signal to clutter ratio for the imaging of mine-like targets achieveable within the considered system-design. As one learns from a glance at Tab. 2. even for dry sand the results do not look too promising. Of course, one cannot discuss the identification capability solely on the basis of the global signal to clutter ratio. Preknowledge on the mine shapes would help to discriminate clutter against a real mine target if the resolution power of the imaging is good enough. Unfortunately, a resolution of 5 cm seems not sufficient to support pattern recognition methods in mine identification.

metallic anti-tank mine	1.26
big plastic mine	0.22
square metallic mine	0.21

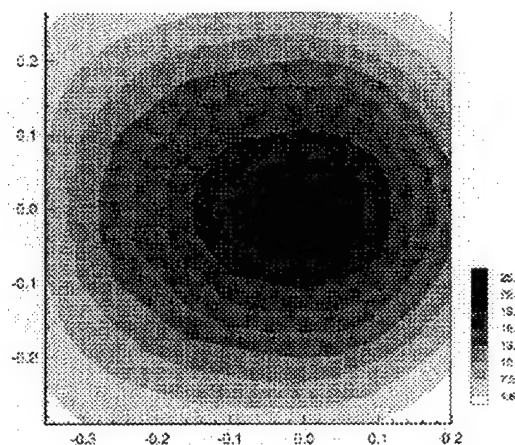Tab. 2: Signal to clutter ratio for targets buried in dry sand under a surface with sharp relief structures

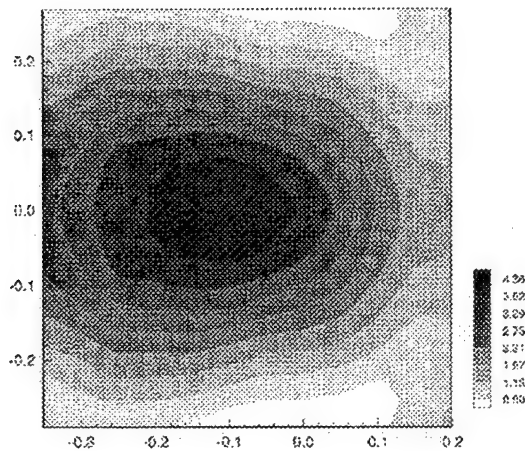Fig. 16: Image produced by a metallic cylindrical mine ⌀ 24 cm, height 8 cm, buried in dry sand

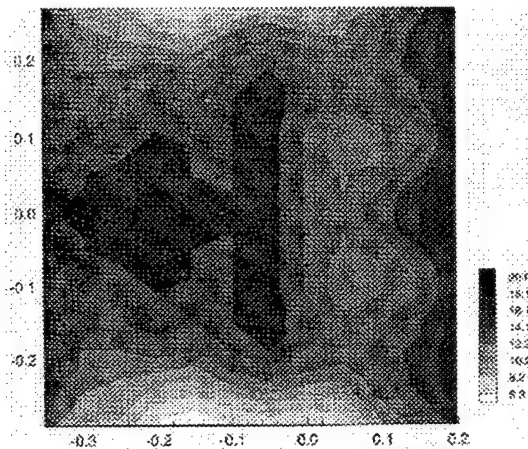Fig. 17: Image produced by a dielectric cylindrical mine ⌀ 24 cm, height 8 cm, buried in dry sand

Fig. 18: Image produced by a rough terrain surface in dry sand
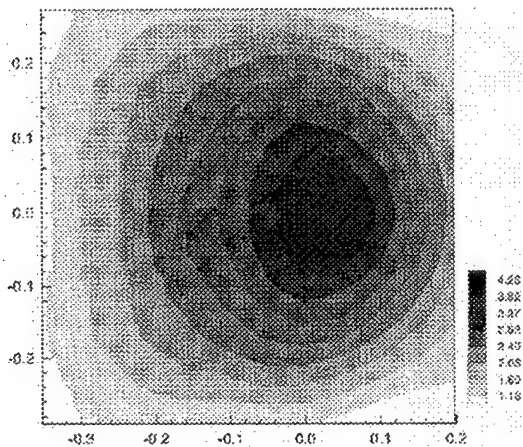
Fig. 19: Image produced by a metallic 10 cm square mine, height 4 cm, buried in dry sand.

5. Conclusions

The efficiency of an electomagnetic imaging principle applied to buried mine searching and identifaction is analysed without worrying much about details of its technical realisation. The usable frequencies for the technique lay in the S-band around 3.5 GHz. This implies that the spatial resolution cannot be much finer than 5 cm. Numerical simulation shows that for favourable soil conditions, e.g. dry sand with a smooth surface, big metallic anti-tank mines and even big plastic mines will actually produce reasonable images with the expected resolution. Pronounced sharp relief-type surface structures, however, may blur the mine images such that only big metallic anti-tank mines could be identified safely under a 10 cm dry sand overburden. In the case of wet clay, mine identification seems to become almost impossible. Attempts to overcome the poor signal to clutter ratio by pattern recognition techniques are hampered by the coarse spatial resolution of the images.

References

1. Habiger K.W., Clifford J.R., Miller R.B., McCullough W.F. (1991) Explosives Detection with Energetic Photons, Nuc. Instrum. and Methods B56/57, 834-838

2. Baum C.E. (1994) Concerning the Identification of Buried Dielectric Targets, Interaction Note 504

3. Graham W.J. (1991) Focused Synthetic Microwave Array for Mine Detection and Imaging, Belvoir Research Report TR-91-001-01

4. Magg M. (1995) Elektromagnetische Feldeinwirkung auf erdüberdeckte Minen, IABG- Report B-TR-M066

5. The MAFIA Collaboration (1994) MAFIA The ECAD System - Manual Version 3.20, CST GmbH, Darmstadt Germany,

Superresolution and Multiresolution SAR/ISAR Imaging

Eric K. Walton
The Ohio State University
Electrical Engineering Department; ElectroScience Laboratory
1320 Kinnear Road; Columbus, Ohio 43212-1191
614 292 7981

· 1. SUMMARY

This paper will discuss radar target imaging using model based spectral analysis applied to synthetic aperture radar (SAR) data or inverse synthetic aperture radar (ISAR) data. The techniques are based on the realization that radar target SAR/ISAR scattering can be modeled as autoregressive. This permits an autoregressive radar scattering model, where the parameters of the model permit model-based imaging.

Finally, we will show a number of different example images where the model based resolution greatly exceeds the resolution available from Fourier based techniques.

2. INTRODUCTION

Wide band radar scattering from a target is often used to form an image of the target. The image is usually referred to as a synthetic aperture radar image (SAR image) or an inverse synthetic aperture radar image (ISAR image) because the transformation behaves as if a very large aperture radar antenna is synthesized from a set of incremental data taken over the domain of the (synthetic) aperture. The SAR image is formed by moving the radar system while the target remains stationary, and the ISAR image is formed by holding the radar at a fixed location while the orientation angle of the target changes.

Conceptually, the transformation involves first transforming the frequency domain data to the time domain (and thus the down range domain) using an inverse spectral estimation usually based on an inverse Fourier transform. Then the complex down range data for each down range value is transformed to the Doppler domain using another spectral estimation, also usually a Fourier transformation [48]. For a rotating target, some points on the target are moving toward the radar (positive Doppler) and some are moving away (negative Doppler). The magnitude of the speed that each point on the target (at each particular down range increment) is moving relative to the radar is proportional to the cross range location of the point. Thus the Doppler domain is the cross range domain. When the down range and cross range data are combined, we have the image domain of the target. Note that there are many implementations of this concept. Some use the two independent Fourier transformations as described (possibly in reverse order),

while others use a 2-D version of the Fourier transform. If the increments are (or can be made) uniform and equal (often including a transformation to Cartesian frequency space, f_x and f_y), then the fast Fourier (or inverse Fourier) algorithm (FFT or IFFT) can be used [29]. (We will ignore backprojection and tomographic-type transformations for now.)

To clarify, a schematic representation of the transformation from a measurement (raw) data set to an image is shown in Figure 1. The raw data is shown in the upper left, where complex radar scattering voltages have been measured as a function of both angle and frequency. If we take an inverse (fast) Fourier transform of each row, we can obtain the data set in the upper right, where each row (and thus each aspect angle) is now a time domain (or down range domain) complex impulse response. (We may call this a waterfall plot of the target.) If we take a (fast) Fourier transform of each column of this data set, then we produce a cross range transformation for each down range column. The data set in the lower right is thus a down range versus cross range image. Each element is complex, but we often display this image by taking the magnitude of each image element and plotting it using a gray scale or color scale proportional to the value of the magnitude.

Note that we may also form the image by taking a (fast) Fourier transform of each column of the raw data set. This forms the Doppler (or cross range) value for each frequency as shown in the lower left corner of the plot. We then take an inverse (fast) Fourier transform of each row to obtain the image domain data set in the lower right.

Finally, note that it is possible to transform directly from the upper left to the lower right by using a two dimensional (fast) Fourier transform (the image must be flipped left to right to correct for the need for an inverse Fourier transform from the left side to the right side of the figure.

It is important to realize that all of these transformations are linear and thus reversible. This is emphasized in figure 1 by the two-way arrows between domains. Many types of image processing techniques involving image domain gating and transformations back to the

measurement domain. The autoregressive techniques discussed later are non-linear, however, and thus not reversible in all cases. Although the reader should keep these reversing concepts in mind, a discussion of these reversible transformations and the impact with respect to superresolution techniques is beyond the scope of this paper and will not be specifically discussed here.

One requirement for such imaging algorithms is that there must be sufficient information about the trajectory of the radar or the target so that the aspect angle is well known. Phase variations due to motion other than simple rotation (e.g. translation) must be removed by compensation based on the known trajectory (this is called motion compensation). Also, the frequency increment and the angle increment must be small enough so that unambiguous (not aliased) transformations to the image domain can be done.

Finally, the frequency band and the angle span must be large enough to provide range and cross range resolution that is meaningful for the target of interest. That means that the range or cross range increment must be small (in some sense) with respect to the extent of the target. This paper is about the case where the last criterion is not met. Either the frequency band or the angle span (or both) is too small to yield sufficient resolution in the image.

There are a number of situations that can limit the frequency band. First, a simple component limitation (antennas, signal sources, receiver bandwidth) is possible for the radar. Alternatively, the propagation medium may be limiting the resolution. (Ground penetration, building penetration, foliage penetration and plasma penetration radars suffer from this limitation.) Finally, the radar target itself may limit the useful observational radar band. It may be a narrow band resonant structure such as a cavity, an active device or a periodic structure.

There are also a number of things that can limit the aspect angle span. The radar platform may not be able to move in an optimum trajectory for SAR, or in the case of an uncontrolled ISAR target, the target simply may not move in such a way as to yield sufficient angle span. In many cases for ISAR, the target is an aircraft or a ship, and it will be undergoing radial motion, pitch, roll and yaw as well as non-uniform angular rotation. These uncontrolled motions limit the useful regions of the angle span to those where the effects are small, or known. Finally, it is also important to observe that the useful angle span may be limited by the degradation in the Doppler versus cross range linear relationship approximation for large aspect angle domains.

It can be seen that the basic steps in the imaging transformations require spectral estimation. So far, we have only mentioned Fourier transformations. The requirements of limited angle or frequency span imaging, however, require that we develop techniques for spectral ·esolution that exceed that available with the Fourier

transform. A large class of model-based modern spectral estimation techniques have recently been developed [54,55], and these so called superresolution transformations in the context of SAR/ISAR imaging will be discussed next.

3. MODEL BASED SPECTRAL ESTIMATION IN THE SAR/ISAR CONTEXT

Superresolution SAR/ISAR imaging techniques are based on an important radar target scattering characteristic. That is that in most radar scattering situations, the radar target behaves as if it were a relatively small number of point scatterers. This means that as the frequency is scanned over the observation band of frequencies, each subcomponent scatterer scatterers a signal that is constant in amplitude and has a phase associated only with its range delay. Since the phase of the scattering from each subcomponent is determined by the round-trip radar range delay distance in wavelengths, a linear scan in frequency results in a linear sweep in radar wavelength and thus a linear phase variation as a function of frequency. We can see that as the radar frequency is scanned, the subcomponent signal is constant in amplitude and varies linearly in phase. It has a complex sinusoidal variation as a function of frequency.

As a function of aspect angle, each subcomponent appears to be moving in a circle. Over a "small enough" sector of angles, the motion can be approximated as a straight line. The apparent speed of each subcomponent depends on the distance from the center of rotation. Thus the subcomponent scatterers appear to have a linear phase variation as a function of aspect angle, with the phase change rate proportional to cross range distance. Thus, the cross range distance is proportional to the (spatial) spectral frequency of the subcomponent scatterer.

The overall target appears to the radar as a relatively small number of such subcomponent scatterers and thus the total received radar signal is a finite linear sum of such terms. The overall signal is a linear sum of a relatively small number of terms, each one of which is a complex sinusoid as a function of frequency.

We have thus moved from the general situation where there is no restriction on the behavior of the scattered signal to the situation where we have a model that says we have a relatively small number of complex sinusoids as a function of frequency and angle. We can use this model to restrict the class of possible solutions for the scattering behavior and thus use a very powerful set of algorithms that take advantage of the a-priori knowledge we have of the radar scattering behavior as a function of frequency and angle.

We start out by modeling the scattered signal as a sum of complex sinusoids [56,63]. Thus in the frequency domain (at each fixed aspect angle), we have

$$\tilde{S}_n = \sum_k A_k \cdot \exp(-j\frac{2\pi \cdot d_k}{c}f_n) + N_n \qquad (1)$$

where

A_k is the amplitude of the kth sub-scatterer

d_k is the down-range distance between the phase zero reference and the kth sub-scatterer

f_n is the nth frequency

N_n is the noise associated with the data at the nth frequency

c is the speed of light

and

$\tilde{S}_n$ is the total complex scattering from the target at the nth frequency.

Now we make a very important realization. We note that any finite sum of sinusoids is autoregressive. This means that the nth data point of the sequence can be accurately predicted by the previous (or future) N data points (N relatively small). Thus for a sum of L complex sinusoids, we have an autoregressive process of order L. In general, we can write

$$x_n = -\sum_{i=1}^{L} a_i \cdot x_{n-i} + u_n \qquad (2)$$

(forward linear prediction (FLP) equation of order L)

where

a_i is the ith prediction coefficient

x_{n-i} is the $(n-i)^{th}$ term in the data series

u_n is the noise (or error) in the prediction for term n

and

x_n is the predicted value in the series.

Similarly, we also have the backward linear prediction (BLP) equation

$$x_n = -\sum_{i=1}^{L} a_i \cdot x^*_{n+i} + u_n \qquad (3)$$

where * indicates the complex conjugate.

Note that the a_i's are unchanged. We note that in general, there is a similar form for the forward-backward linear prediction (FBLP) equation. This takes advantage of the unchanged nature of the a_i's.

Once we are able to determine the coefficients of these prediction equations, we can obtain a power spectral estimation of the original data series by a z-transform. In this example, where the original data is a set of scattering measures as a function of frequency, this represents a transformation from the frequency domain to the time (or down range) domain.

$$h(t) = \frac{1}{1 - \sum_{k=1}^{p} a_k \cdot \exp(-j\pi \cdot 2t/T)}, t = -\frac{T}{2}, \cdots, \frac{T}{2} \qquad (4)$$

where $h(t)$ is the power profile (actually the square of the impulse response). Note that with no loss of generality a_1 can be set equal to one. Examination of this equation shows that there is no inherent limit in the resolution of the time domain estimate. This is due to the fact that the model has perfect knowledge of the signal once the parameters of the model have been estimated. Of course, if the parameters are poorly estimated, the model has perfect knowledge of the wrong system and thus we will have a very high resolution estimate of the wrong value for the spectral estimate. Never forget, we must distinguish between resolution and accuracy!

Note that the expression in (4) is an evaluation of (2 or 3) along the unit circle in the complex plane. As a result, the location of the poles and zeros of (4) determine the behavior of (2 or 3).

The entire field of model based spectral estimation depends on the process used to find the coefficients that must be inserted in the z-transform (4). This process can be understood for FLP by expanding equation (2) into a matrix equation ($u_n = 0$).

$$\begin{bmatrix} x_{p-1} & x_{p-2} & \cdots & x_1 \\ x_p & x_{p-1} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & \cdots & \cdots & x_{-p-1} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = -\begin{bmatrix} x_p \\ x_{p+1} \\ \vdots \\ x_N \end{bmatrix} \qquad (5)$$

This matrix representation of the forward linear prediction (FLP) equation helps to show that the equation simply represents an overdetermined set of linear equations. There are $N+1$ equations in p unknowns where the size of the measurement set is $N+1$ and the number of sinusoids is p. The solutions for the a's thus involve methods of finding efficient and effective solutions to this set of equations.

Similar representations are also available for the backward linear prediction (BLP) equations and the forward-backward linear equations (FBLP) [34].

4. MATRIX SOLUTIONS

Using either the FLP or the FBLP approach thus involves solving a set of overdetermined linear equations based on a set of measured values. There are two main methods used to solve these equations. The first is the least squares method and the second is the total least squares method [54,55].

The underlying assumption of a least square solution is that the errors in the set of equations are imbedded in the

prediction column. In fact, if the elements of the matrix are noisy, the column space of the matrix is only an approximation of the solution space for the equations. (The total least squares method is used to improve on this assumption.)

The matrix (5) equation can be written in general as

$$Y \cdot a = g \, . \tag{6}$$

If the matrix Y represents data values of a perfect autoregressive (AR) process, the matrix would have rank p. In practice, the matrix is full rank because there is always some noise present. As a result, all rows of the matrix are linearly independent and there is no exact solution to the set of equations.

The consequence of there being no exact solution to the set of equations is that the vector g is not in the vector space that is spanned by the columns of the matrix Y. The nature of the least squares solution is to consider a vector $g+r$, where r is a small perturbation vector. If r is chosen so that $g+r$ is in the column space of Y so that r is minimized, the associated a is called the Least Squares solution of $Ya=g$.

In more geometric terms, the vector g is projected onto the column space of Y and the vector r is the perpendicular error vector between the column space and the g vector. The solution vector then represents the linear combination of the column vectors in the column space that exactly equal the projections.

It is known from linear algebra that the projection of the vector g onto the column space of Y can be represented as $Y^H g$ (where H represents the Hermetian transpose). Therefore, multiplying both sides of $Ya=g$ by Y^H yields

$$Y^H Y a = Y^H g \tag{7}$$

It follows that the least squares solution of (6) can be represented as

$$a = (Y^H Y)^{-1} Y^H g \tag{8}$$

The inverse of $Y^H Y$ is known to exist because Y is full rank.

And there we have it; a least Squares solution to the overdetermined set of equations given in (5).

Note that the underlying assumptions of a Least Squares solution are that the errors are in the g vector. That is , the elements of the matrix Y are perfectly known. This is the assumption made when the g vector is projected onto the column space of Y. In fact, if the elements of the matrix are noisy, which they are in practice, the column space of the matrix is only an approximation of the solution space for the equation. The Total Least Squares method is then used to improve on this assumption,

especially in the case where the elements of the matrix Y are assumed to have noise associated with them. A detailed discussion of the TLS method is beyond the scope of this paper, and can be found elsewhere.

5. THE BURG METHOD

A third method of finding the model parameters of a FBLP model is called the Burg Method [54,56]. Burg said that the parameters should be selected in such a way as to maximize the entropy of the signal outside the measurement window. (In fact, the method is often called the Maximum Entropy Method.) Using this approach, and a modified version of the Levinson algorithm, the model parameters can be found. The details of the Burg algorithm are found in the references.

6. APPLICATIONS OF SUPERRESOLUTION TRANSFORMATIONS

A number of specific imaging examples will next be used to demonstrate these techniques, moving from one-dimensional imaging to two dimensional imaging.

6.1 Angle Domain Estimation Example

Radar scattering measurements are often done in a compact RCS measurement range [57]. This system uses a large reflector in the near field of the test area to transform a spherical wave from a small feed to a plane wave in the test area. Problems arise from spurious scatterers in the neighborhood of the reflector that produce spherical (near-field) signals in the test zone. Locating the source of such error terms can be done by measuring the field received in the test zone and then transforming to a direction of arrival map to trace signals back to their origin. This is especially difficult in this situation because the error signal is a spherical (or cylindrical) wave and thus if one attempts to increase the resolution of the direction of arrival estimate by using a larger aperture scan, the spherical nature of the signal causes defocusing.

As a study of this type of problem, vertical scan was taken in the Ohio State University (OSU) compact RCS measurement range, and small increments of displacements were used to compute the direction of arrival [41]. One result is shown in figure 2a. Note that the response due to the main beam gives a large signal lobe at zero aspect, but the smaller error terms are not seen. If we take larger spans over the test zone, the previously mentioned defocusing also hides the effect due to the spurious error signal. If an autoregressive model-based spectral estimation technique is used to compute the direction of arrival spectrum at a set of vertical increments in the test zone, we can obtain the plot shown in figure 2b. A low order autoregressive model was used here. Note that the spurious signal is now visible. The direction of arrival varies depending on the location of the scan in the test zone. This can be made meaningful, if directional lines are plotted from each sub-scan region in the test area back toward the compact range main reflector as shown in figure 3. Note that the

lines converge at the junction where the parabolic curvature of the main reflector transitions into a cylindrical skirt. We can estimate the amplitude of the error signal with respect to the main beam signal for each directional line, and these are shown on the plot.

This is a specific example where model based spectral estimation is very effective at extraction of spurious signals in an environment where it would otherwise not be possible.

6.2 Range domain estimation

The previous example showed that model based spectral estimation techniques can be used to estimate direction of arrival by processing spatial domain data. It is also possible to use these techniques to transform from the frequency domain to the time and thus down-range domain [28,34,53,56]. An example case is shown in figure 4. This example was done for a generic missile shape consisting of a cylinder with a frustum (nose cone) and rounded edge wings and a vertical blade. The overall length was 153 cm. Data over the frequency band from 5 to 6 GHz will be shown here. Now, since we have only 1 GHz of radar bandwidth, our range resolution (-3dB main time domain response width) using Fourier transforms is 1 ns. (the inverse of the bandwidth). An example down range profile is shown in figure 4 as a dashed line. If we use an autoregressive algorithm applied to the set of complex frequency domain values to estimate the down range profile, we can obtain the solid line. This line can now show (1) the tip of the frustum, (2) the frustum-cylinder junction, (3) the leading edge of the wings and blade, (4) the wing and blade tips, (5) the trailing edge of the blade, (6) the cylinder end, the discontinuity and (7) the across the rear caustic term. Remember that one of the most important characteristics of these model based techniques is not the increase in time accuracy of a single scattering term, but the ability to distinguish between two closely space scatterers such as found at the end of this cylinder.

We can form a backprojection or tomographic image of this target if we sum up the contribution of each down range profile across the image domain at each aspect angle. For the Fourier based profiles (equation 4), we have a signed (positive and negative voltage contributions) summation where small error terms will average out in regions where there is no target. This result (only over 180° of observation domain) is shown in figure 5a. For the autoregressive down range profiles, it is the nature of the spectral estimation that only power estimates are available. This means that small error terms will add coherently, rather than integrating to zero. On the other hand, we note that there are in fact, no "noise" terms as such in the down range profiles. Each profile is deterministic (although the terms may have errors) and thus there is no random noise type behavior in the no-target regions of the image. Thus when the tomographic summation over the target image domain is done (also

180° only), it is possible to obtain the image shown in figure 5b. Note that the vertical blade, the wing tips, and both edges of the trailing end of the cylinder can be seen in this image with much greater resolution than the Fourier based image.

6.3 Hybrid Image Domain Estimation

If the measurement data set (such as is shown in the upper left of figure 1) is constrained in only one dimension, then we need only use the model based spectral estimation technique in the transformation along the dimension that is constrained. Two examples, based on this constraint will be given below.

6.3.1 *Small aircraft target*

Consider the example from a x-band radar imaging test series on a small private aircraft as shown in figure 6.[39,31,28,20]. In this radar test, the radar was fixed in location and operating with a radar wavelength of approximately 3 cm. The radar operated from 9150 to 9898.8 MHz in 128 steps of 5.85 MHz. The small propeller driven aircraft was flying in a small diameter circle while several miles from the radar. Frequency scans were taken every 25.6 ms. For this set of data, 128 such scans corresponds to approximately 4.9 degrees azimuth change. The raw data was first phase shifted in such a way as to remove the radial component of velocity from the data. The result is a data set that behaves as if the aircraft is fixed in location, but rotating about a point. (The process is called motion compensation.) The image shown in figure 6a is the result of a two- dimensional Fourier transform using only 32 data points from the angle span. The full frequency set was used (thus no special limitation in the down range domain). The data were cosine windowed and zero padded to 128 points in both down range and cross range. Note the image of the aircraft (flying away from the radar in a turn to the left). Note also the diagonal lines near the top of the image. This is propeller modulation effects producing false image terms (an image alias). The image shown in figure 6b was produced by first transforming to the down-range domain using an inverse fast Fourier transform (IFFT) (top left to top right in figure 1). This effect has the benefit of reducing the number of scattering terms in each subsequent transform because there is only a small number of scattering terms at each down range location. The upper right data matrix (shown in figure 1) was then transformed column-by-column to the image domain by using an FBLP spectral transform with an autoregressive order of only 4. (I.E.: it allowed at most 4 scatterers in each down-range cut.) Note the highly resolved image of the aircraft near the center of the image. Also note the propeller modulation terms manifested as diagonal lines near the top of the image.

This entire process can be done again, but with only 8 aspect angles included. The FFT-based image for this case is shown in figure 7a while the FBLP based image (also order 4) is shown in figure 7b. Note that the image of the aircraft has significantly degraded in the FFT

based image, but that the basic aircraft size and shape can still be seen as well as the propeller modulation terms in the FBLP based image.

6.3.2 Target large in wavelengths
One of the criticisms of these model based spectral estimation techniques is that the number of spectral terms must be relatively small for the technique to be successful (only relatively low model order will work). We include an example of 3 cm wavelength imaging of a 122 meter long ship to demonstrate the capabilities of the hybrid image in this application [1, 6].

In this exercise, we obtained RCS scattering data for a ship that was moving in a 1 mile diameter circle approximately 3 miles out to sea. The radar was on the shore. There was a limitation in the ability to image this ship using Fourier techniques because the pitch, roll and yaw effects tended to dominate and defocus the image if data were collected over a time long enough to permit significant motion in those dimensions. On the other hand, if only a short data sequence was used, the aspect angle span was too short for useful cross range resolution in the image. Motion compensation was also an issue and angular perturbations caused estimation of the angular velocity of the vessel to be in error.

One of the FFT based images is shown in the upper left plot in figure 8 for this ship. The down range transform used the full 9.10 to 9.228 GHz band and thus realized a down range resolution of 1.2 meters. The angle domain span used to form the image was only from 156 to 158.7°, however, and this limited the cross range resolution so that the image of the ship was quite "smeared out" in cross range. The result of the hybrid model based spectral estimation for cross range imaging (model order only 7) is shown in the right hand plots of this figure. (Note the scaled drawing of the ship in the lower left hand portion of the figure.) With proper motion compensation applied, the overall length and shape of the ship can be seen in the image.

6.4 Data Extrapolation Imaging
One other approach to superresolution SAR/ISAR imaging will be discussed here. Note from equations 2 and 3 that these are prediction equations (the coefficients are called predictor coefficients) in the sense that elements of the measurement data set, (the x_n values in the equation) are predicted based on a model based linear extrapolation. The prediction coefficients, a_i, are derived based on the errors between the predicted values and the actual measurement values. Note that once the values of the predictor coefficients are available, however, it <u>is</u> possible to make a model based extrapolation of the data outside the measurement data set. In the diagram of the SAR/ISAR process given in figure 1, this means that if one of the dimensions of the measured data set (either the frequency band or the angle span) is not large enough to permit the image resolution

required, then the a_i's can be used to extrapolate outside the measurement space. Once this has been done, the typical Fourier based imaging algorithm can be applied to the expanded data set [22,29,30].

The process, thus is to extend the experimental measurement data set (the upper left matrix in figure 1) and then form the image using Fourier based techniques.

One limitation of this technique is that there may be measurement errors in the original data set that cause inaccuracies in the a_i's. This will lead to degradation of the predicted values as the extension distance increases.

Also, this is a model based technique, and if the physical situation is such that the data do not fit the model (especially outside the measurement domain), then the prediction will be inaccurate as the domain of the prediction increases. As an example, note that the phase progression (the distance between the radar and the target subcomponent) does not change linearly as a function of aspect angle. This means that the cross range behavior of the model based extrapolation will not be effective in predicting the values beyond the measurement angle span. To avoid this problem, one can "polar format" the scattered field data and can then extrapolate the polar formatted data along the two Cartesian components of frequency that result (see [28,29,34]). (Note that this property is not always a disadvantage. We are attempting, after all, to increase the spectral resolution in the observed data set only in most cases.)

An example where this has been done is given in figure 9(taken from [29]). The example data set is an air to air VHF/UHF ISAR measurement set on a KC135 jet [39]. The frequency band was originally measured from 160 to 360 MHz and the angle span from -20 to 20°. For the analysis shown here, we used only the frequency band from 240 to 280 MHz and the angle data only from -2 to 2°. Figure 9a shows the FFT based ISAR image.

Next, FBLP autoregressive techniques are used to compute the prediction coefficients and extend the available frequency band and angle span by a factor of approximately 3. In this case, a 2-D FBLP algorithm of order 5 was used [34] to obtain the prediction coefficients. The algorithm used the full 2-D autoregressive nature of the raw ISAR data in Cartesian space (f_x and f_y) [29].

Although there is no limit to the extrapolation extent, the studies reported here suggest that extending the domain by a factor of three is a reasonable limit. Note that the data will be predicted over the entire measurement domain. In the method presented here, only the data outside the measurement domain is derived from the prediction coefficients, the measurement data itself is retained where available. Finally, the total span of (predicted and measured) data is windowed (e.g., Hamming). The windowing reduces the transform

domain sidelobes while de-emphasizing the energy of the extrapolated data. Finally, Fourier transforms are used to transform to the image domain. Note that one may choose to extrapolate the frequency band and/or the angle span (although this extrapolation is actually applied in 2-D to the Cartesian frequency components derived from the polar format transformation).

The resulting image of the KC135 aircraft is shown in figure 9b (taken from [29]). Note the tail and wing tips as well as the four jet engine outlets in the image. (Remember that the data were collected from the rear.) Clearly, the resolution of the image is much higher than that of the image in figure 9a.

7. CONCLUSION

This paper has discussed radar target imaging techniques (SAR and ISAR) based on model based autoregressive spectral estimation. We have discussed the basics of autoregressive spectral estimation and have given examples of the resulting images as compared to Fourier based imaging techniques.

We began with spatial domain extrapolation, then showed a tomographic image technique where the image was formed from the high resolution down range power profile. Using an outline of general SAR/ISAR concepts based on figure 1, we then showed a hybrid algorithm, where only one of the image domains was extended, and a full 2-D FBLP autoregressive technique where both domains (in Cartesian frequency space) were extended.

In all cases, it was demonstrated that with model based spectral estimation it is possible to obtain resolution in the SAR/ISAR image well in excess of that available from Fourier techniques.

8. REFERENCES

This is mostly a list of papers that have gone into the overall development by the author and close associates of the techniques presented here. The references by Kay [54] and Marple [55] are included as excellent modern spectral estimation reference books. The reference for Mensa [48] is included as an excellent reference on high resolution radar imaging.

1. Cai, L., I. Gupta and E. K. Walton, "Inverse Synthetic Aperture Imaging Studies of a Ship at X-band." Antenna Measurement Techniques Association 17th Meeting and Symposium, Williamsburg, VA, November 13-17, 1995.

2. Gerry M. and E. K. Walton, "Low Frequency Radar Target Imaging," Antenna Measurement Techniques Association 17th Meeting and Symposium, Williamsburg, VA, November 13-17, 1995.

3. Moghaddar A. and E. K. Walton, "Region-of-Interest Tomography using Biological Image Perception Concepts," 1995 International Symposium on Signals, Systems and Electronics, San Francisco, October 26, 1995.

4. Tu, M-W., I. J. Gupta and E. K. Walton Radar Image Processing using Efficient Maximum Likelihood Estimator, Progress Report 727721-6, The Ohio State University ElectroScience Laboratory, August 1995.

5. Dewi, L., I. J. Gupta and E. K. Walton, Polargram Radar Imaging Technique, Progress Report 728243-3, The Ohio State University ElectroScience Laboratory, February 1995.

6. Walton, E. K. and Gupta, Final Report on Radar Superresolution Processing Development Final Report 728243-4, The Ohio State University ElectroScience Laboratory, February 1995.

7. Walton, E. K., "The Measurement of Radar Cross Section," Chapter 5 of Radar Target Imaging Edited by W.-M. Boerner and H. Uberall, Springer-Verlag, Berlin, 1994. Volume 13 of the Springer Series on Wave Phenomena. ISBN 3-540-57791-2/ SAP-No. 10062915.

8. Walton E. K. and M. J. Gerry, "Analysis of dispersive radar scattering by data preconditioning," in Ultra-Wideband, Short-Pulse Electromagnetics," ed: L. Carin and L. B. Felsen, Plenum Press, New Your, NY 1994. (Second International Conference on Ultra-Wideband Electromagnetics)

9. Tseng, H.-W. and E. K. Walton, "Experimental RCS Analysis of a Communications Antenna Mounted on a Large Cylinder," 1994 Antenna Measurement Techniques Assn. Meeting, Long Beach, CA, October 3-7, 1994.

10. Moghaddar, A., Y. Ogawa and E. K. Walton, "Estimating the Time-Delay and Frequency-Decay Parameter of Scattering Components using a Modified MUSIC Algorithm," IEEE Trans. on Antennas and Propagation, Vol. 42, No. 10, pp. 1412-1419, October, 1994.

11. Gerry, M. and E. K. Walton "Analysis of Amplitude Dispersion in Radar Scattering using Preconditioned Linear Prediction," 1994 Antenna Measurement Techniques Assn. Meeting, Long Beach, CA, October 3-7, 1994.

12. Gupta, I., E. K. Walton and M.-W. Tu, "Application of ML Estimation To Radar Imaging," 1994 Antenna Measurement Techniques Assn. Meeting, Long Beach, CA, October 3-7, 1994.

13. Onendahal, I., E. K. Walton and I. Gupta, "Enhanced High Resolution Radar Imaging," Antenna Measurement Techniques Assn. Meeting, Long Beach, CA, October 3-7, 1994.

14. Tu, M.W., I.J. Gupta and E.K. Walton Application of ML Estimation to Radar Imaging, Technical Report 727721-4, The Ohio State University ElectroScience Laboratory, August 1994.

15. Walton, E. K. and I. J. Gupta, Characterization of Hughes Aircraft Company Small (Chamber 1) Compact Range, Technical Report 727721-3, The Ohio State University ElectroScience Laboratory, July 1994.

16. E. K. Walton and M. J. Gerry, "Analysis of dispersive radar scattering mechanisms from measured data," AP-S/URSI meeting, June 19-24, 1994.

17. Walton, E. K. and F. Paynter, "Comparison of Impulse and Noise-Based UWB Ground Penetrating Radars," URSI Radio Science Meeting, Seattle, WA, June 19-24, 1994.

18. Gerry M. and E. K. Walton, "Analysis of Dispersive Radar Scattering by Data Preconditioning," URSI Radio Science Meeting, Seattle, WA, June 19-24, 1994.

19. Moghaddar and E. K. Walton, "A Data-Adaptive Time-Frequency Representation applied to the Scattering from a Jet Aircraft," IEEE AP-S Int. Symposium, Seattle, WA, June 19-24, 1994.

20. Moghaddar, A. and E. K. Walton, Description of the Computer Program EM-TFD (Time-Frequency Distribution of Electromagnetic Scattering), Technical Report, The Ohio State University ElectroScience Laboratory, June 1994.

21. Young, J. D., E. K. Walton, J. S. Gwynne and M. J. Gerry, Superresolution Image Processing Final Report 727149-6, The Ohio State University ElectroScience Laboratory, May 1994.

22. Odendaal, J. W., I. J. Gupta and E. K. Walton, Enhanced Radar Imaging, Tech. Rpt. 728243-1, The Ohio State University ElectroScience Laboratory, March 1994.

23. Moghaddar A., and Eric K. Walton, "Time-Frequency Distribution Analysis of Frequency Dispersive Targets," in Ultra-Wideband, Short Pulse Electromagnetics Edited by H. Bertoni et al., Plenum Press, 1993.

24. Walton, E. K., A. Jain, C. R. Boerman, V. J. Vokurka, "Hughes Aircraft Company RCS/Antenna Measurement Chamber Characterization," 1993 Antenna Measurement Techniques Assn. Meeting in Dallas, October, 1993

25. Walton, E. K., A. Moghaddar and Y. Ogawa, "Superresolution Analysis of Frequency Dispersive Scattering," 1993 Antenna Measurement Techniques Assn. Meeting in Dallas, October, 1993

26. Walton, E. K., "Modern Radar Signal Processing Algorithms Applied to Ground Penetrating Radar," Second Government Conference On Ground Penetrating Radar, Columbus, Ohio October, 1993.

27. Walton, E. K. "Ground Penetrating Radar using Ultra-Wideband Noise," Second Government Conference On Ground Penetrating Radar, Columbus, Ohio October, 1993.

28. Walton, E. K., and I. J. Gupta, "Superresolution ISAR Imaging Techniques," 1993 Antenna Measurement Techniques Assn. Meeting, Dallas, October 1993.

29. J. Gupta and M. J. Beals, "High Resolution Radar Imaging using Data Extrapolation," 1993 AMTA Meeting, Dallas, Texas, October, 1993

30. Beals, M. J. and I. J. Gupta, Radar Imaging using Data Extrapolation, , Tech. Report 727721-1, The Ohio State University ElectroScience Lab., Columbus, Ohio September, 1993.

31. Walton, E. K. and M. J. Gerry, "High Resolution Time Domain Analysis of VHF/UHF RCS Measurements," IEEE Antennas and Propagation Soc. Meeting, Ann Arbor, July 1993.

32. Gerry, M. J. and E. K. Walton, Behavior of Modern Spectral Estimation Algorithms for Transformation of Radar Scattering Data From the Frequency Domain to the Time Domain, Technical Progress Report 727149-3, The Ohio State University ElectroScience Laboratory, June 1993

33. Walton, E. K. and Huan-Wan Tsieng Development of Theoretical and Experimental RCS Analysis for a Communications Antenna, Technical Progress Report 312606-2, The Ohio State University ElectroScience Laboratory, June 1993.

34. Walton, E. K., I. J. Gupta, M. W. Tu and A. Moghaddar Modern Spectral Estimation Techniques for Radar Target Imaging, Annual Report 312587-2, The Ohio State University ElectroScience Laboratory, May 1993

35. Moghaddar A. and Eric K. Walton, "Time-Frequency Distribution Analysis of Scattering from Waveguide Cavities," IEEE Trans. on Antennas and Propagation, Vol. 41, No. 5, pp. 677-680, May 1993.

36. Walton, E. K. and Huan-Wan Tsieng Development of Theoretical and Experimental RCS Analysis for a Communications Antenna, Technical Progress Report 312606-1, The Ohio State University ElectroScience Laboratory, February 1993.

37. Walton, E. K. and H. W. Tsieng, Development of Theoretical and Experimental RCS Analysis for a Communications Antenna, Technical Progress Report 312606-3, The Ohio State University ElectroScience Laboratory, February 1993

38. Moghaddar and I. J. Gupta, "High Resolution Radar Imagery using Parametric Modeling and Data Extrapolation," Proceedings of the 1992 AMTA Meeting, Columbus, OH October, 1993

39. Jain, A. and I. Patel, "ISAR imaging of aircraft-in-flight using ground based and airborne radars," Proceedings of the 1992 AMTA Meeting, Columbus, OH October, 1993.

40. Tu, M.-W. and E. K. Walton Inverse Synthetic Aperture Radar Imaging using a Hybrid Superresolution Algorithm, Tech. Rpt. 312587-1, The Ohio State University ElectroScience Laboratory, Department of Electrical Engineering, August 1992.

41. Walton E. K. and A. Moghaddar, "Imaging of a Compact Range Using Autoregressive Spectral Estimation," IEEE Aero. & Elect. Sys. Magazine, Vol. 6, No 7, pp. 15-20, July 1991.

42. Walton E. K. and I. Jouny, "Target Identification Using Bispectral Analysis of Ultra-Wideband Radar Data," in Ultra-Wideband Radar: Proceedings of the First Los Alamos Symposium , Ed: Bruce Noel CRC Press, Boca Raton, July 1991.

43. Jouny, I. and E. K. Walton, "Applications of the Bispectrum in Radar Signatures Analysis and Target Identification," International Signal Processing workshop on Higher Order Statistics, Vale, Co, July, 1991.

44. Jouny, I. and E. K. Walton,"Target Identification in Non-Interactive Clutter using the Bispectrum," IEEE Antennas and Propagation Society; URSI Symposium, June, 1991.

45. DeMattio, C. and E. K. Walton , Target Identification of Actual Aircraft using Features Derived from the Impulse Response and the Bispectrum, Technical Report 529819-9, The Ohio State University ElectroScience Laboratory, June 1991.

46. DeMattio, C. and E. K. Walton , Target Identification of Actual Aircraft using Features Derived from the Impulse Response and the Bispectrum, Technical Report 529819-9, The Ohio State University ElectroScience Laboratory, June 1991.

47. Jouny, I, F. D. Garber, R. L. Moses and E. K. Walton," Applications of the Bispectrum in Radar Signature Analysis and Target Identification," Proceedings of the 1991 Automatic Object Recognition Session, Society of Photo Optical Instrumentation Engineers Meeting, April, 1991

48. Mensa, D. L., High Resolution Radar Cross Section Imaging, Artech House, Dedham, Mass., 1990.

49. Jouny, I. E. K. Walton, R. L. Moses and F. D. Garber, Bispectral Analysis of Radar Signals with Application to Target Classification, Technical Report 723090-2, The Ohio State University ElectroScience Laboratory, August, 1990.

50. Jouny, I. and E. K. Walton, "Bispectral Analysis of Radar Signals," Joint Microwave Theory and Techniques Society; IEEE Antennas and Propagation Society and URSI Symposium, May, 1990.

51. Walton E. K. and I. Jouny, "Bispectrum of Radar Signatures and Application to Target Classification," Radio Science, Vol. 25, No. 2, pp. 101-113, March-April 1990.

52. Jouny, I. and E. K. Walton, "Target Identification Using Bispectral Analysis of UWB Radar Data," First Los Alamos Symposium on Ultra-Wideband Radar, March, 1990.

53. Walton, E. K., "Far Field Measurements and Maximum Entropy analysis of Lossy Material on a Conducting Plate," IEEE Trans. on Ant. & Prop., Vol. AP-37, No. 8, August 1989.

54. Kay, S. M., Modern Spectral Estimation: Theory and Application, Prentice Hall, Englewood Cliffs, NJ, 1988.

55. Marple, S. L. Jr., Digital spectral analysis with Application, Prentice-Hall, Englewood Cliffs, NJ, 1987.

56. Walton, E. K., "Comparison of Fourier and Maximum Entropy Techniques for High-Resolution Scattering Studies," Radio Science, Vol. 22, No. 3, pp. 350-356, May-June 1987.

57. Walton, E. K. and J. D. Young, "The Ohio State University Compact Radar Cross Section Measurement Range", IEEE Trans. On Antennas and Propagation, Vol. AP-32 (11), pp. 1218-1223, November 1984
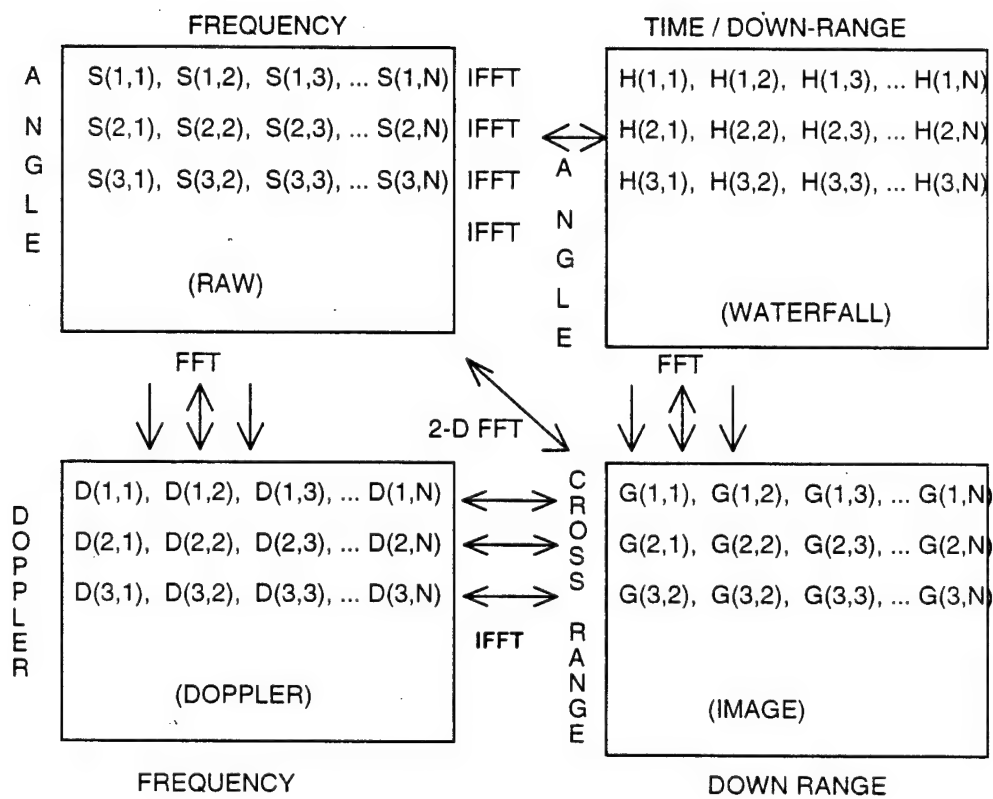
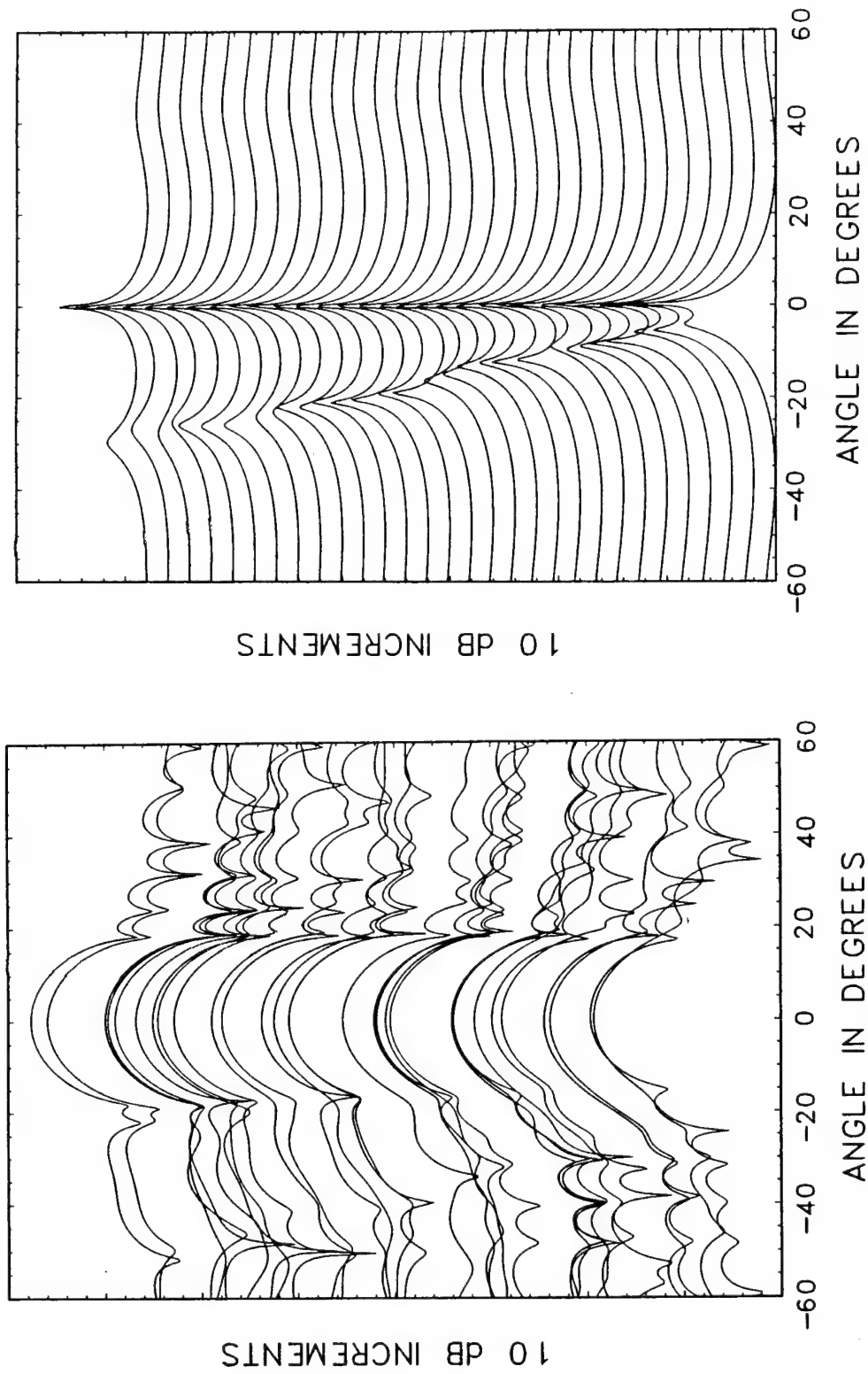Figure 1. Schematic diagram of the SAR/ISAR algorithm.

Figure 2. Angular spectrum for a set of 26 cm. sub-apertures with 6.3 cm spacing (8 GHz). (a) Fourier transformation; (b) Autoregressive spectral estimation. (from [41])
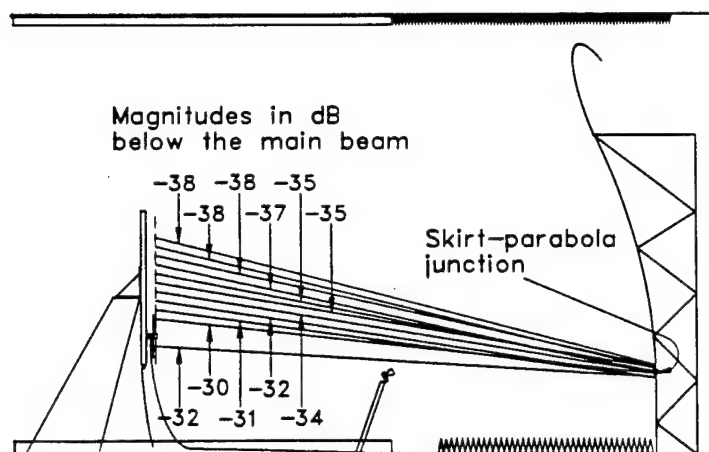
Figure 3. Direction of arrival and relative (to main beam) amplitude of secondary signal for a set of sub-apertures from figure 2.b. (from [41])
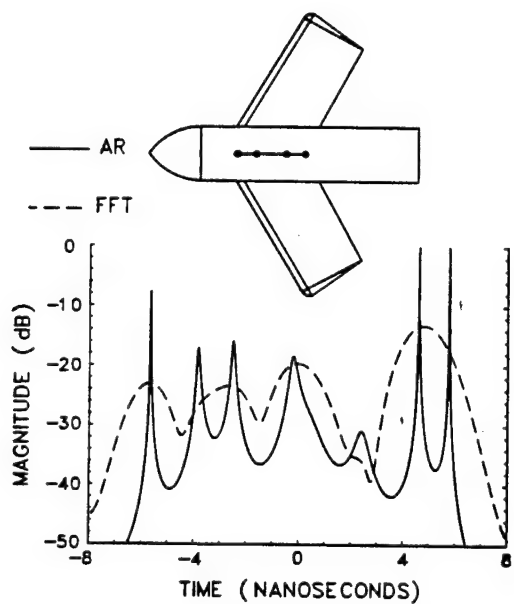

Figure 4. Time domain amplitude response for generic missile shape over the band 5 to 6 GHz. (dashed line = Fourier results; solid line = autoregressive spectral estimation.
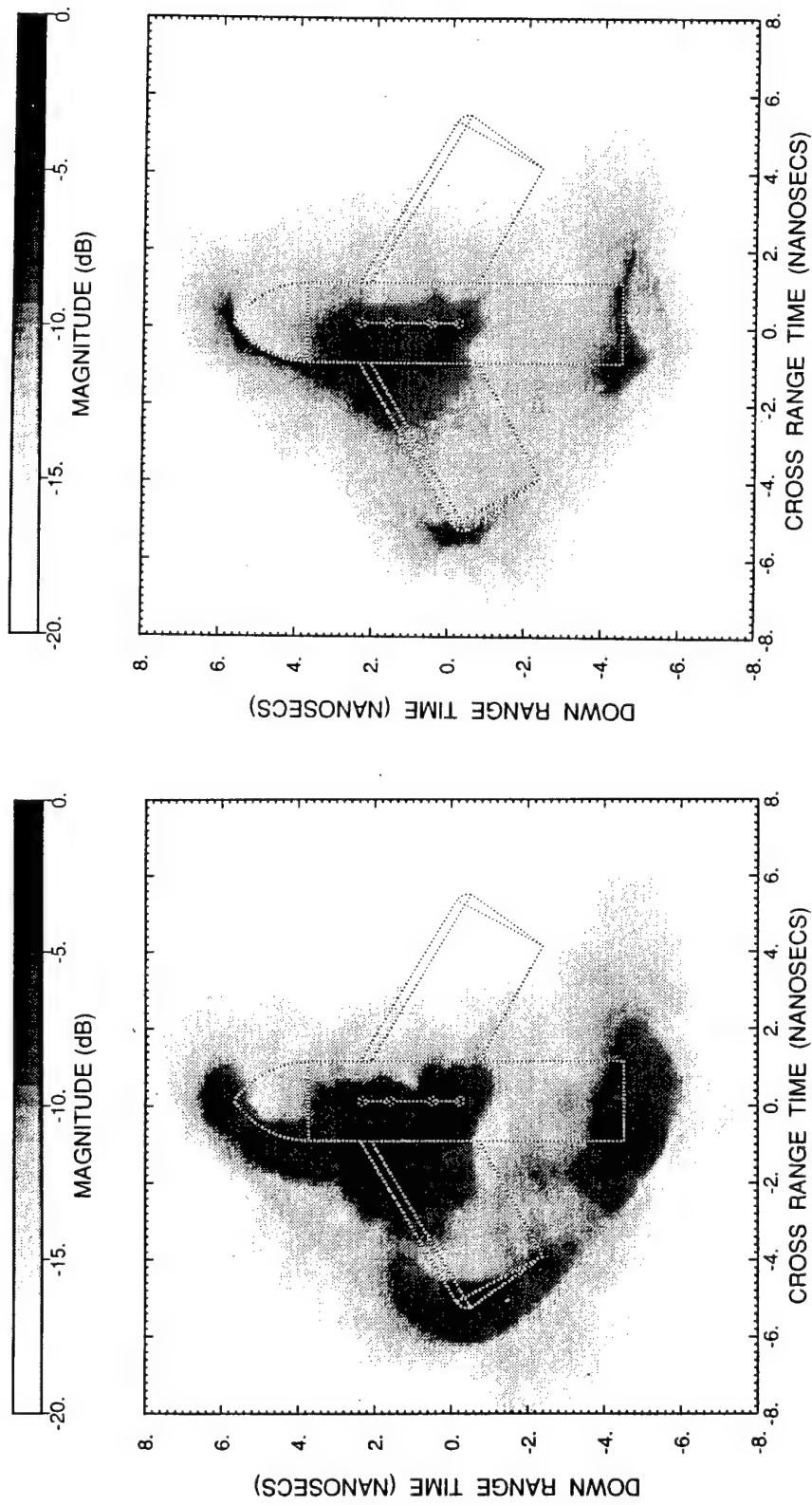
Figure 5. Backprojection ISAR image of generic missile shape over the band 5 to 6 GHGz. (a) Fourier transforms; (b) Autoregressive spectral estimation techniques.
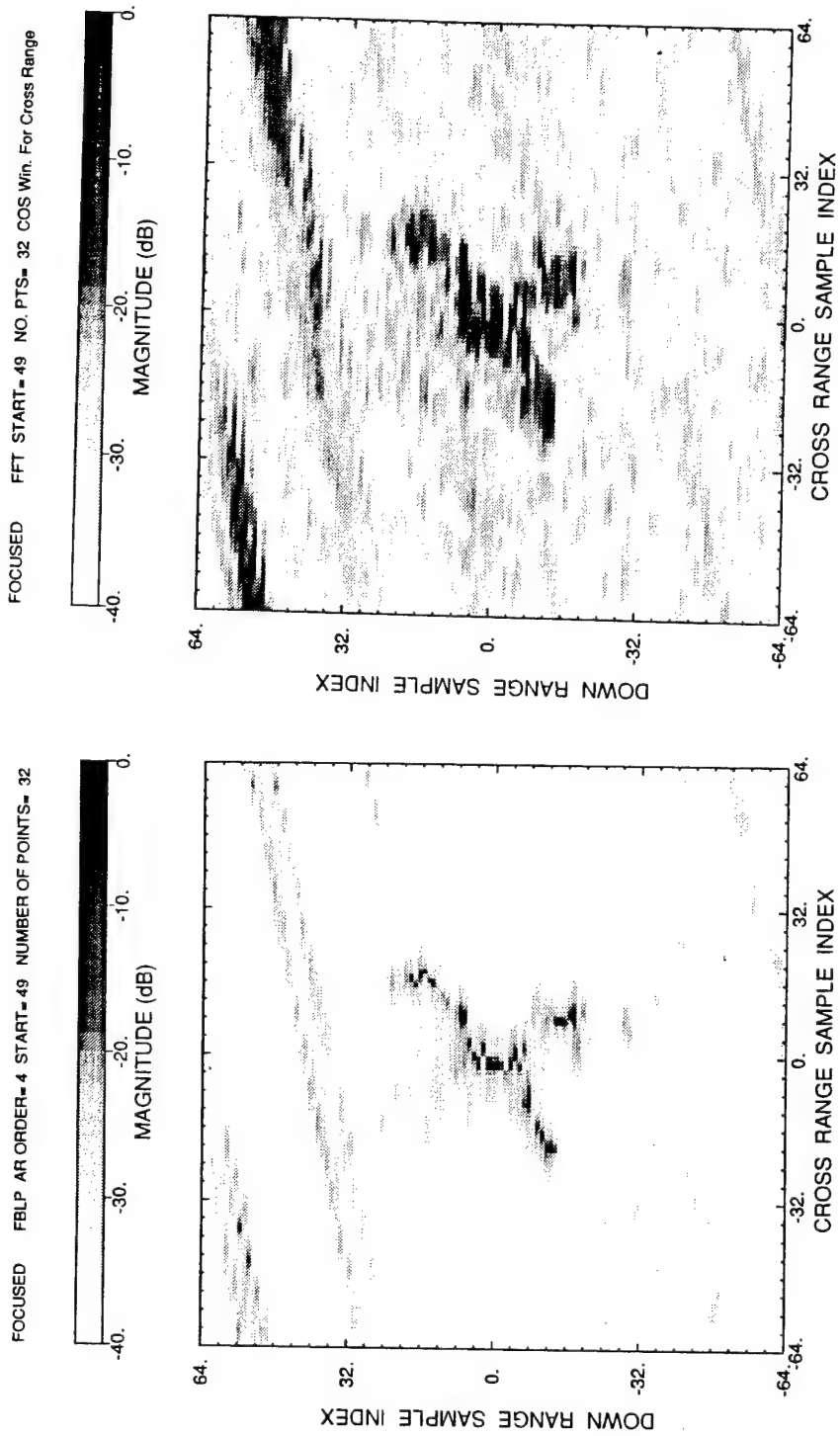
Figure 6. ISAR image of small aircraft (Mooney 231) (32 points in angle domain) (a) FFT - with cosine window; (b) Hybrid FBLP image.
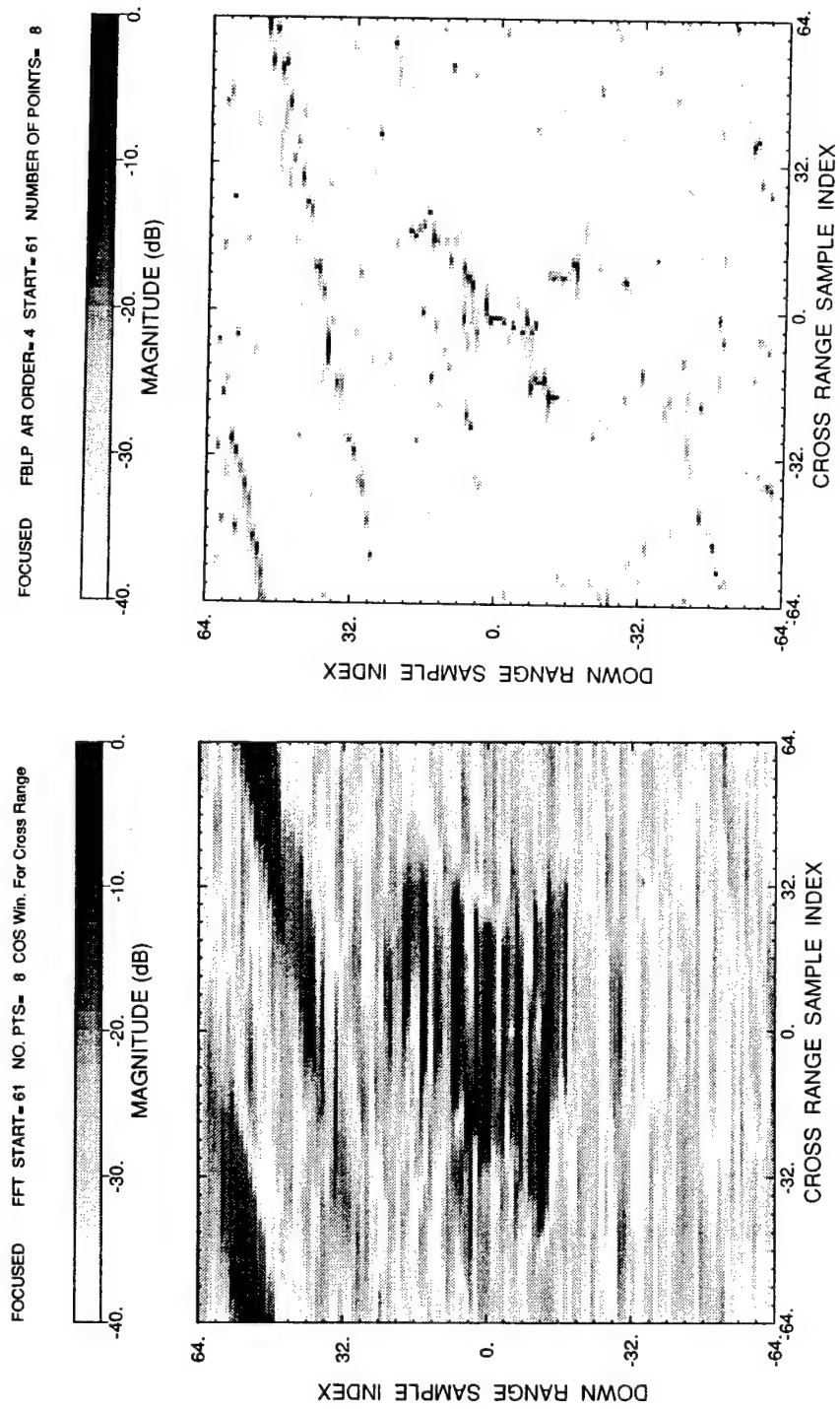
Figure 7. ISAR image of small aircraft (Mooney 231) (8 points in angle domain) (a) FFT - with cosine window; (b) Hybrid FBLP image.
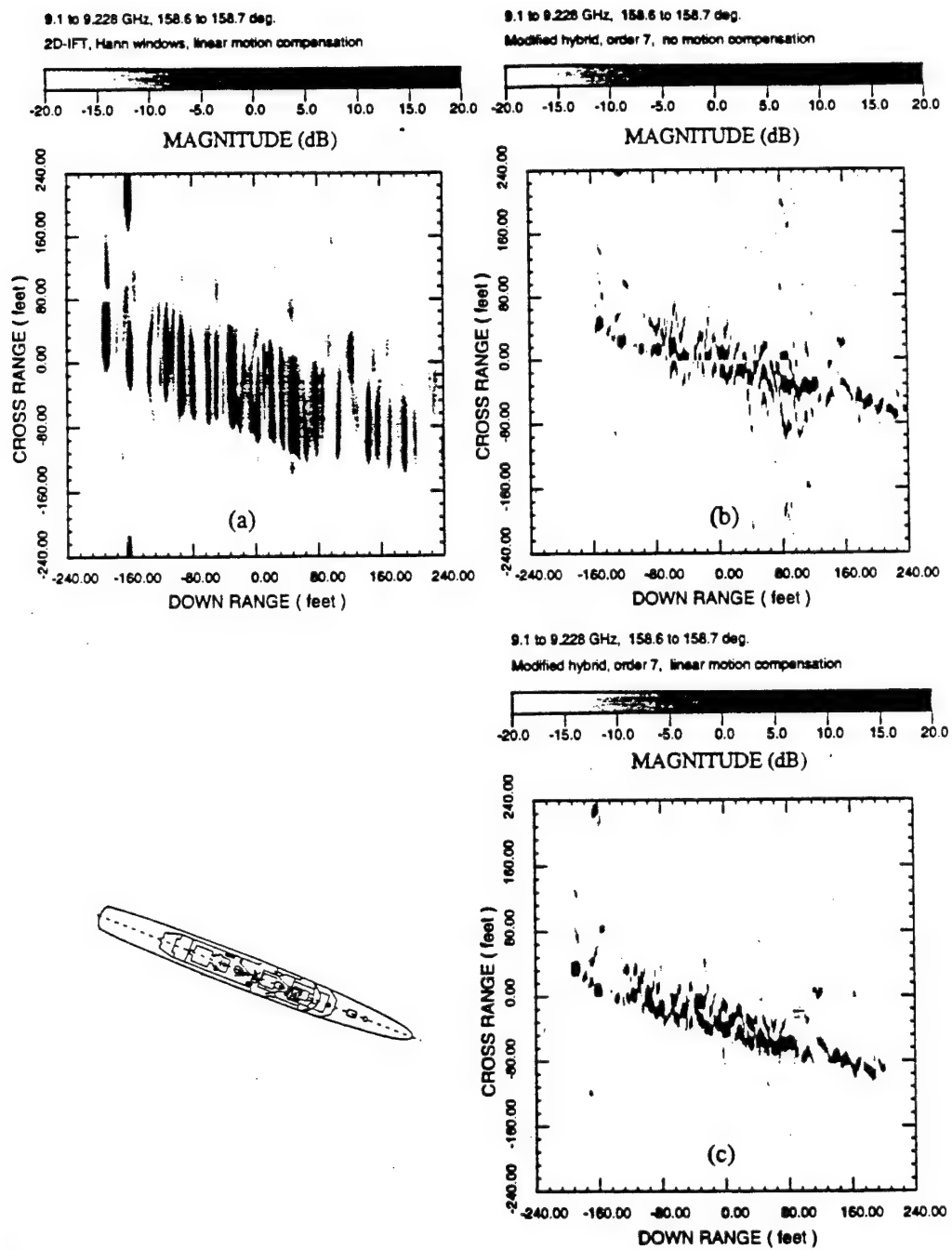
Figure 8. ISAR image of naval ship. 9.100 to 9.228 GHz; 158.6 to 158.7 angular degree span. (a) = Hanning windowed 2-D IFFT; (b) Modified hybrid technique of order 7 (in cross range) without motion compensation; (c) Modified hybrid technique of order 7 (in cross range) linear motion compensation. Note scaled ship image in lower left. (from [1])

Original Data, .24 to .28 GHz; -2,2 Deg. 2-D FBLP Order=5

MAGNITUDE (dB)

CROSS RANGE (FT)

DOWN RANGE (FT)

Original data .24 to .28 Ghz, -2,2 degrees; No Extrapolation

MAGNITUDE (dB)

CROSS RANGE (FT)

DOWN RANGE (FT)
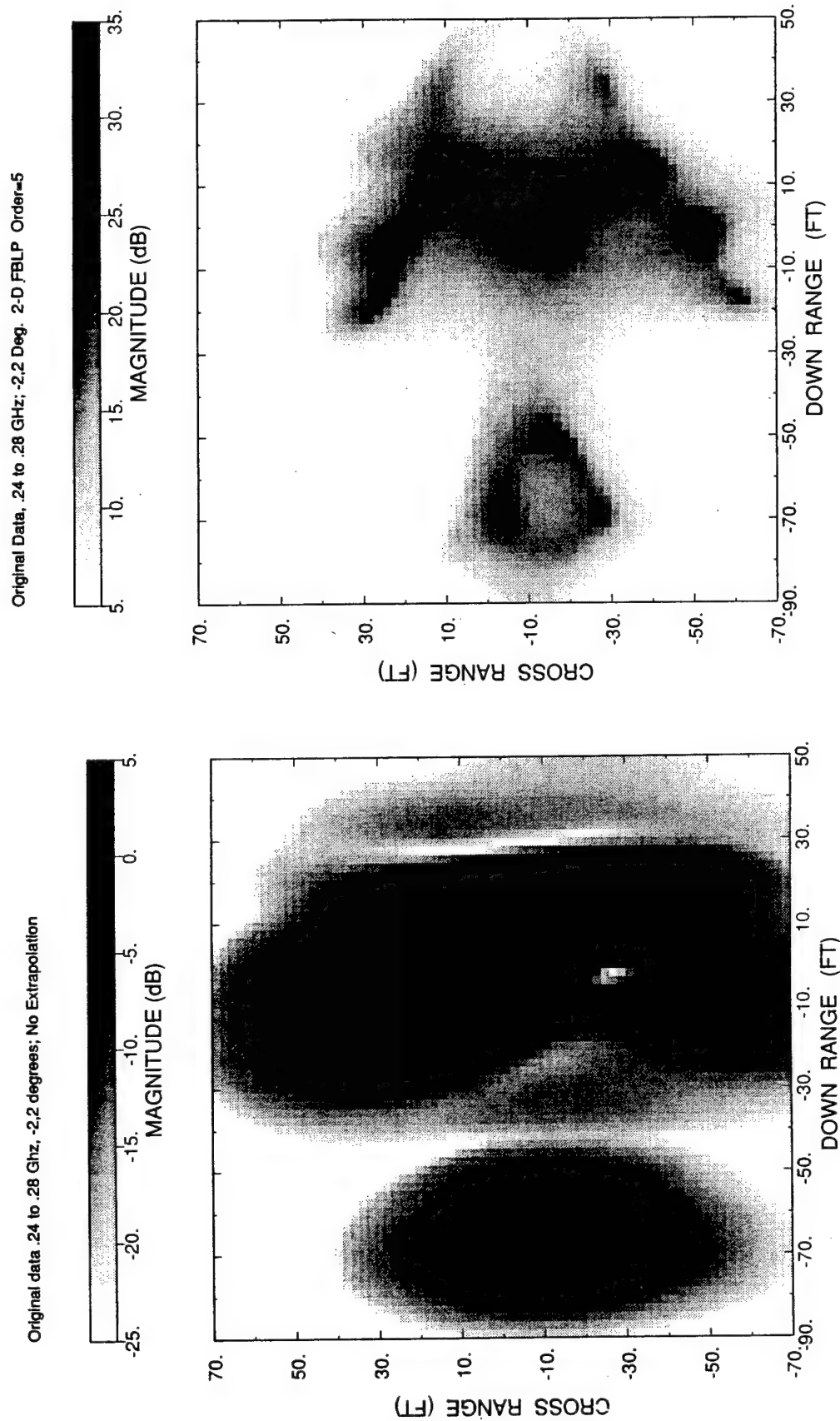
Figure 9. ISAR image of KC-135 aircraft (240 to 280 MHz; -2 to 2 degrees off rear aspect) (a) 2-D FFT; (b) Spectral and angular extrapolation by factor of 3 using 2-D FBLP of order 5. (from [28])

Recognition of Buried Targets

Carl E. Baum
Air Force Research Laboratory
AFRL/DEHP, Building 909
3550 Aberdeen Avenue S. E.
Kirtland AFB, NM 87117-5776
U. S. A.

1. SUMMARY

A difficult problem concerns the identification of buried mines and unexploded ordnance (UXO). This paper reviews the use of the singularity expansion method and symmetry to identify such targets and distinguish them from other buried objects of comparable size.

2. INTRODUCTION

In trying to locate and identify UXO and mines in soil, one is faced with a formidable problem. One needs to discriminate these from rocks, tree roots, shrapnel, etc. Otherwise one will have a large false-alarm rate. Here we emphasize electromagnetic techniques utilizing target signatures. (See C. E. Baum, Signature Based Target Recognition, in this volume.) For small targets in soil adequate resolution in imaging can be very difficult due to high-frequency attenuation and scattering in the medium. For large targets (such as buried waste dumps) imaging may be more appropriate, but this is not the case considered here.

An important target signature in this case is the natural frequencies in the singularity expansion method (SEM). This is manifest in more than one form.

3. ELECTROMAGNETIC SINGULARITY IDENTIFICATION (EMSI)

In this case, a ground penetrating radar (GPR) illuminates the target with a pulse (or many frequencies) with significant frequencies in the general range of 100 MHz to 1 GHz so that wavelengths in the external medium are of the order of the target dimensions, thereby exciting the resonances. The target natural frequencies are significantly changed from their values in free space by the properties of the surrounding medium (soil,

water). For metal targets there are scaling relationships for the natural frequencies, modes, and coupling coefficients based on the constitutive parameters of the external medium. For dielectric targets one can use the internal resonances with a perturbation formula if the wave impedance of the exterior medium is lower than that of the dielectric target.

4. MAGNETIC SINGULARITY IDENTIFICATION (MSI)

For highly, but not perfectly, conducting targets (metal) there is a very promising technique based on a set of natural frequencies corresponding to diffusion in the target. The frequencies are quite low (kHz range) so that wavelengths in the external medium are large compared to distances to the exciting and measuring antennas (typically coils in the near field). Provided the soil is non magnetic (permeability approximately that of free space), the natural frequencies are independent of the external medium, a significant advantage. These natural frequencies are on the negative real axis of the s plane (Laplace-transform variable or complex frequency) and correspond to decaying exponentials in time domain. Together with appropriate unit vectors these natural frequencies can be used to characterize the magnetic polarizability dyadic, and thereby the target.

Besides buried targets, this technique is being applied to security systems such as airport metal detectors.

5. ACOUSTIC SINGULARITY IDENTIFICATION (ASI)

Acoustic/elastodynamic scattering also has a singularity expansion representation with similar properties. However, in a medium such as water

which only supports p (pressure) waves with no significant s (shear) waves, there is a lack of polarization information, a disadvantage. Its great advantage applies in cases of external media with a conductivity large enough to severely attenuate electromagnetic waves at frequencies of interest.

6. SYMMETRY IDENTIFICATION
In conjunction with the above techniques, one can use target symmetry as an identifier, both for EMSI and MSI. In the case of EMSI a target with the symmetry of a body of revolution with axial symmetry planes, and with the rotation axis perpendicular to the ground surface has no hv scattering

(cross polarization) in the usual h,v radar coordinates. This has been observed with synthetic aperture radar (SAR).

7. CONCLUDING REMARKS
The various techniques discussed above each have advantages and limitations. As such, they are complementary. For the three kinds of natural frequencies one can look at these in a diagrammatic way as in fig. 1. This has the properties of a matrix and a Venn diagram (as in Boolean algebra). Note in the bibliography the various "Notes" are published by AFRL/DEHP, C. E. Baum, editor.
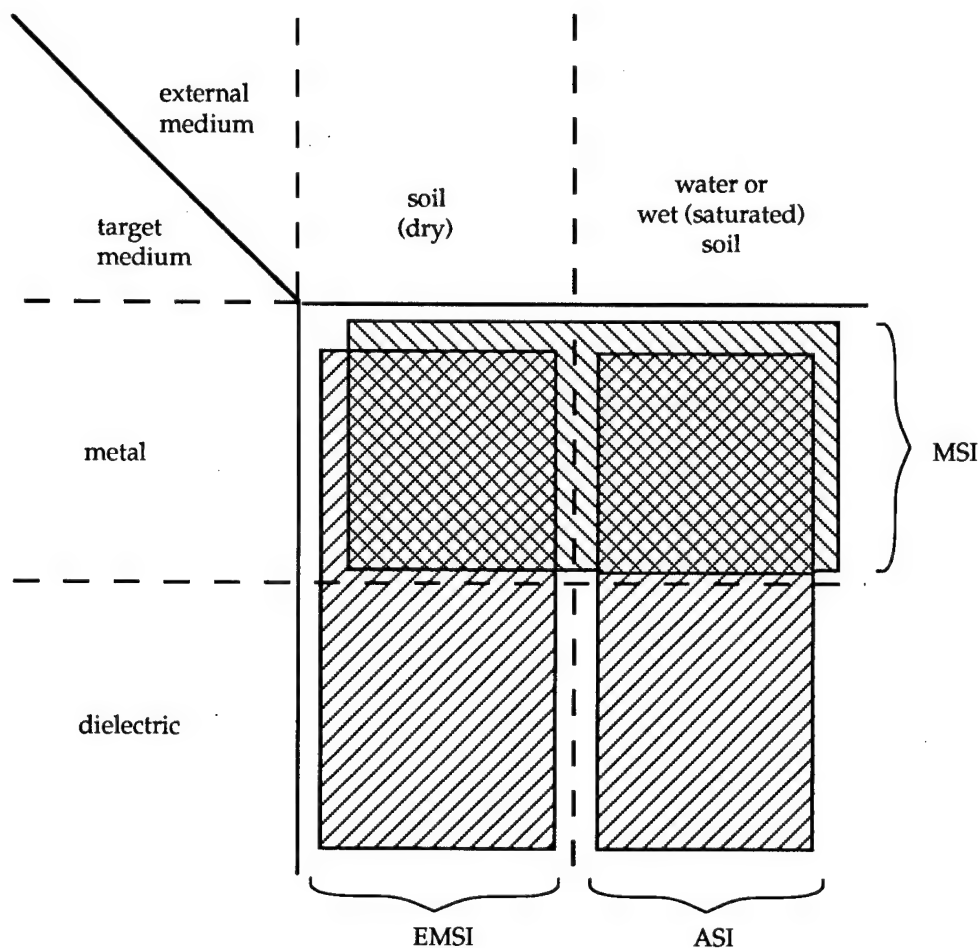
Fig. 1. SEM-Based Buried-Target Identification.

BIBLIOGRAPHY

General

1. C. E. Baum, Discrimination of Buried Targets via the Singularity Expansion, Inverse Problems, Vol. 13, 1997, pp. 557–570.

2. C. E. Baum (ed.), *Detection and Identification of Visually Obscured Targets*, Taylor & Francis (in publication).

Electromagnetic Singularity Identification (EMSI)

General

3. L. Peters, Jr., and J. D. Young, Applications of Subsurface Transient Radar, ch. 9, pp. 296–351, in E. K. Miller (ed.), *Time-Domain Measurements in Electromagnetics*, Van Nostrand Reinhold, 1986.

4. L. Peters, Jr., J. J. Daniels, and J. D. Young, Ground Penetrating Radar as a Subsurface Environmental Sensing Tool, Proc. IEEE, 1994, pp. 1802–1822.

"Perfectly conducting" targets

5. C. E. Baum, The SEM Representation of Scattering from Perfectly conducting Targets in Simple Lossy Media, Interaction Note 492, April 1993.

6. S. Vitebskiy and L. Carin, Resonances of Perfectly Conducting Wires and Bodies of Revolution Buried in a Lossy Dispersive Half-Space, IEEE Trans. Antennas and Propagation, 1996, pp. 1575–1583.

7. S. Vikebskiy, L. Carin, M. A. Ressler, and F. H. Le, Ultra-Wideband, Short-Pulse Ground-Penetrating Radar: Simulation and Measurement, IEEE Trans. Geoscience and Remote Sensing, 1997, pp. 762–772.

8. G. W. Hanson and C. E. Baum, Perturbation Formula for the National Frequencies of an Object in the Presence of a Layered Medium, Interaction Note 532, October 1997.

9. C.-C. Chen and L. Peters, Jr., Buried Unexploded Ordnance Identification via Complex Natural Resonances, IEEE Trans. Antennas and Propagation, pp. 1645–1654.

Dielectric targets

10. C. E. Baum, Concerning the Identification of Buried Dielectric Targets, Interaction Note 504, July 1994.

11. G. W. hanson and C. E. Baum, A Volumetric Eigenmode Expansion Method for Dielectric Bodies, Interaction Note 517, August 1996.

12. G. W. Hanson and C. E. Baum, Perturbation Formula for the Internal Resonances of a Dielectric Object Embedded in a Low-Impedance Medium, Interaction Note 520, August 1996.

13. M. C. Worthy and C. E. Baum, A Comparison of Exact Versus Perturbed Pole Locations of Dielectric Objects in Dielectric Medium, Interaction Note 529.

14. M. C. Worthy and C. E. Baum, A Library of the Natural Frequency Responses for Cylindrical Shaped Buried Plastic Mines, Interaction Note 530, July 1997.

15. M. C. Worthy and C. E. Baum, A Library of the Natural Frequency Responses for Rectangular Shaped Buried Plastic Mines, Interaction Note 531, July 1997.

Magnetic Singularity Identification (MSI)

16. C. E. Baum, Low-Frequency Near-Field Magnetic Scattering from Highly, But Not Perfectly Conducting Targets, Interaction Note 499, November 1993.

17. C. E. Baum, The Magnetic Polarizability Dyadic and Point Symmetry, Interaction Note 502, May 1994.

18. G. D. Sower and S. P. Cave, Detection and Identification of Mines from Natural Magnetic and Electromagnetic Resonances, pp. 1015–1024, in Detection Technologies for Mines and Minelike Targets, Proc. SPIE, Vol. 2496, 1995.

19. C. E. Baum, N. Geng, and L. Carin, Integral Equations and Polarizability for Magnetic Singularity Identification, Interaction Note 524, March 1997.

20. G. D. Sower, Eddy Current Responses of Canonical Metallic Targets: Theory and Measurement, Interaction Note 526, May 1997.

Acoustic Singularity Identification (ASI)

21. G. Bollig and K. J. Langenberg, The Singularity Expansion Method as Applied to the Elastodynamic Scattering Problem, Wave Motion, Vol. 5, 1983, pp. 331–354.

22. H. Überall (ed.), *Acoustic Resonance Scattering*, Gordon and Breach Science Publishers, 1992.

23. C. E. Baum, The SEM Representation of Acoustic and Elastodynamic Scattering, Interaction Note 512, June 1995.

24. G. C. Gaunard and H. C. Strifors, Transient Resonance Scattering and Target Identification, Appl. Mech. Rev., Vol. 50, No. 3, March 1997, pp. 131–148.

Symmetry Identification

25. C. E. Baum, Symmetry in Electromagnetic Scattering as a Target Discriminant, pp. 295–307, in H. Mott and W.-M. Boerner (eds.) Wideband Interferometric Sensing and Imaging Polarimetry, Proc. SPIE, Vol. 3120, 1997.

26. L. Carin, R. Kapoor, and C. E. Baum, Polarimetric SAR Imaging of Buried Landmines, IEEE Trans. Geoscience and Remote Sensing (in publication.)

Mine-detection test facilities at TNO-FEL test location "Waalsdorp"

Jan Rhebergen and Peter Zwamborn
TNO Physics and Electronics Laboratory
PO Box 96864
2509 JG The Hague
The Netherlands

1.   SUMMARY

As part of the TNO-FEL Ultra-Wide-Band Ground-Penetrating-Radar (UWB-GPR) project, a test facility for controlled GPR experiments was planned. Construction of this sand-box test facility has recently been completed. At the same site another test facility, for evaluating various commercial of the shelf (COTS) sensors for the Dutch humanitarian de-mining program HOM2000, has been realised. This article describes the test facilities themselves as well as the framework in which they were realised.

2.   INTRODUCTION

In recent years the research into mine-detection systems at the TNO Physics and Electronics Laboratory (FEL) has increased substantially. Apart from establishing a firm knowledge base and experience through ambitious projects such as UWB-GPR, the TNO-FEL has also been actively involved with the Dutch ministry of defence's, HOM2000, humanitarian demining project. Both the TNO-FEL de-mining/mine-detection focus-program, of which UWB-GPR is a key component, and the involvement in the HOM2000 humanitarian program have spawned the need for good testing and experimenting facilities. Over the past year, two rather uniques facilities have been developed and completed. The first is a facility specifically for testing UWB-GPR systems and for data acquisition. The second is a multi-sensor mine-detection test facility for evaluating systems under controlled but realistic conditions.

3.   TESTING FACILITY FOR CONTROLLED UWB-GPR EXPERIMENTS

In this section we will describe the development of the first test facility which has been constructed at TNO-FEL for carrying out controlled experiments with a ground-penetrating radar system. This test facility has been developed in cooperation with the Delft University of Technology, where a parallel project in the area of ground-penetrating radar is being carried out. One of the principal goals was to perform the neccesarry analytical and design work to support the construction of an experimental facility, where target objects can be buried in a sand medium and illuminated by an ultra-wideband electromagnetic field. The sand medium should be a realistic, yet controlled environment in which objects of interest can be buried. The electromagnetic field scattered by the buried object has to be observed and diagnosed for the benefit of developing an computer database containing the responses of a large number of objects (i.e. landmines etc.)

3.1   COOPERATION WITH DELFT UNIVERSITY OF TECHNOLOGY

The Delft University of Technology, is engaged in a ground-penetrating radar project called "Improved ground-penetrating radar technology". This project started in January 1996 and will end in December 1999. It is supported by the Dutch Technology Foundation STW, (Stichting voor de Technische Wetenschappen) under contract number DMB55.3649. The objective of the project is the development of a new electromagnetic technique for the geological characterisation of the shallow subsurface of the earth. In this particular STW project four different parts are distinguished.

- The first part of the project comprises the design and realization of a transportable ground-penetrating radar system. This subproject concentrates on the development of antennas in a frequency band that ranges from 200 MHz up to 1 GHz. It consists of a transmitting and receiving device, in which time-domain signals (nanosecond pulses) are generated, received and processed.
- The second part of the project contains the construction of a full-size testing site in order to perform controlled experiments with the ground-penetrating radar system.
- The third part of the project comprises the development of a fast and accurate electromagnetic modeling method. This modeling method provides the basis for the electromagnetic imaging and furthermore provides a means to check the performance of a field measurement over a known geological situation. Adequate measurements at the testing site should provide the input data.
- The final part of the project comprises the development of practical strategies and processing algorithms with the final goal, the realization of the actual image. Not only the test site will be used, also more realistic geological sites will be investigated and a compared to existing ground-penetrating tools and conventional techniques.

The development of the new ground-penetrating radar system at Delft University of Technology will take place in a cooperation between the International Research Center for Telecommunications Transmission and Radar (IRCTR) and the Centre for Technical Geoscience (CTG), research school. IRCTR is hosted by the Faculty of Electrical Engineering and is directed by Prof. Dr. L.P. Ligthart. CTG is a cooperation between the Laboratory of Electromagnetic Research of the Faculty of Electrical Engineering (directed by Prof. Dr. P.M. van den Berg) and the section Technical Geophysics of the

Faculty of Mining and Petroleum Engineering (directed by Prof. Dr. J.T. Fokkema). Prof. Dr. J.T. Fokkema is also the project leader of the STW project and as such the point of contact for TNO FEL.

The third part of the STW project (the development of a electromagnetic modeling method) is partly sponsored by TNO-FEL. TNO-FEL has awarded a research grant for a post-doc to participate in the project for one year. The second part of the STW project, which is the construction of the test facility, is performed in cooperation with the Electromagnetics Section of TNO FEL, since such a testing facility was also needed for the another TNO FEL project namely,UWB GPR. The test facility was built on the premisses of TNO FEL (completed July 1997) and will be used for both projects. It is clear that a cooperative effort in the development of such a testing facility has many advantages, which not only includes the sharing of the financial burden.

The project at Delft University of Technology sponsored by STW and the UWB-GPR project at TNO-FEL sponsored by the Netherlands Army are complementary; knowledge and experience that is gained in the projects can be exchanged as well as for example measurement data. Both projects differ in signal processing; the STW project aims at detection only; characterization or classification is postponed until the imaging step can be satisfactorily performed. TNO-FEL aims at the exploitation of the singularity expansion method to perform detection *and* classification. Furthermore, since a university is not a defence research institution, the goal of their ground-penetrating radar system is directed towards civil applications, like engineering activities and environmental and archeological investigations. Examples are the detection of cables and pipes, the determination of the stability of dikes, the detection of leakage and the and the spreading of (in)organic pollution and so on, and not primarily landmines.

3.2 CONSTRUCTION OF THE TESTING FACILITY

Within the framework of the cooperation with the Delft University of Technology, a full size testing facility has been designed and constructed on the premisses of TNO FEL, in order to perform controlled radar experiments. The testing facility has been completed in July 1997. It is located in the dunes directly behind the TNO-FEL building, on a location that was formerly known as the "EMP site" where electromagnetic pulse experiments were once carried out.

The test facility consists of a buried wooden box. The dimensions of this box are 10 m x 10 m wide and 3 m deep (300 cubic meters). The box has been fully dug in into the ground, hence the bottom of the box located three meters below the ground level. Special care has been taken not to use any metal parts in the construction of the box or in the vicinity thereof. The sandbox is filled with clean (homogeneous) river sand from the (Dutch) Caland canal. In order to keep the condition of the sand in the box optimal and to prevent pollutions from the outside (for example ground water) entering the box, a drainage system was installed and the inside of the box was covered with a watertight plastic lining. To prevent the weather from influencing the test conditions and to protect the measuring equipment, a large tent covers the entire site. Additionally, a wooden shed is available for storage of personal computers and measurement equipment. While filling the box with sand, special care was taken to get a homogeneous profile. Current tests indicate that this was not entirely successful but should be good enough to work with. Later on in this study we will probably empty the box and refill the box with a different ground composition. To facilitate the measurement of EM transmissions into the ground a square PVC tube running from the surface of one side to the bottom of the other side has been installed about one meter from the edge of the sandbox.

On the next page we show a small photo gallery which shows the progress in the construction of the UWB-GPR test facility.
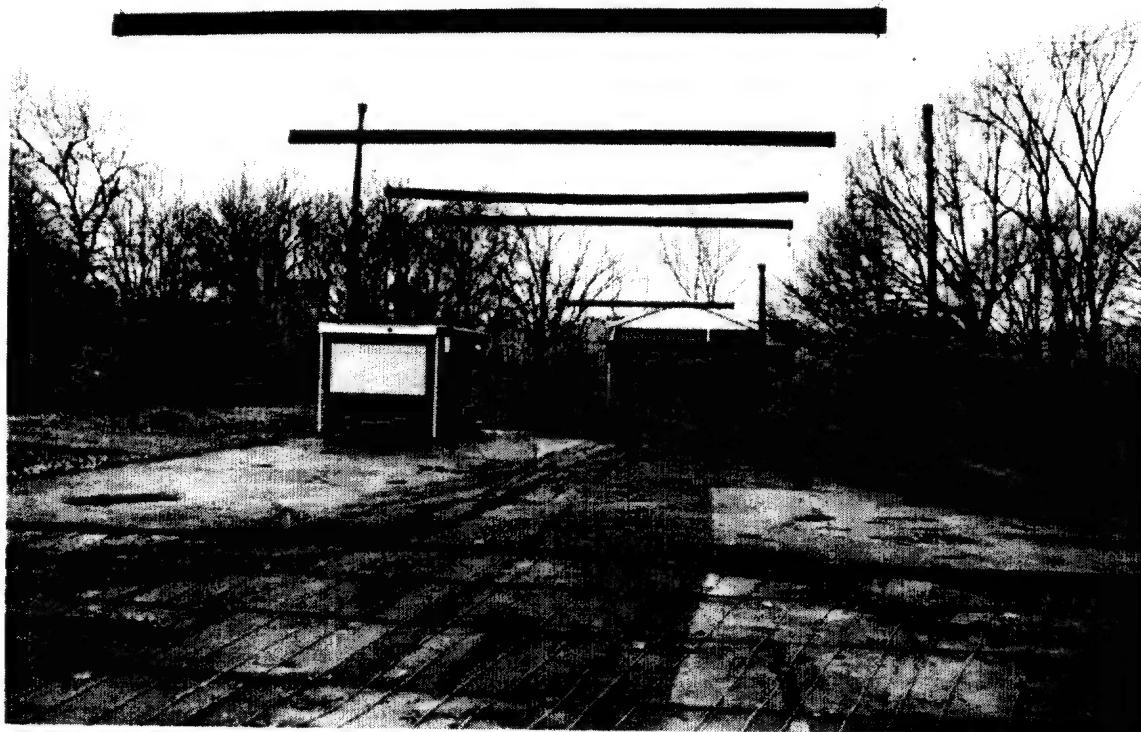
Figure 3.1   Old situation. Location of the test site in the dunes near TNO-FEL. This location was formerly known as "EMP site" where electromagnetic pulse experiments were carried out.

Figure 3.2   Start of the activities. Dismantling of the EMP site. The location is being prepared for construction activities. In the background the (green) wooden storage shed, where the equipment is kept, is visible.

Figure 3.3   Start of the construction of the wooden box. First, some wooden piles are driven into the ground. The piles are used as a framework to strengthen the construction of the walls.

Figure 3.4   Construction of the walls of the box using wooden beams. The wood that was used to construct the box was tropical hardwood from Central Africa.

Figure 3.5   While the walls of the box were being constructed, the soil at the location was removed simultaneously.
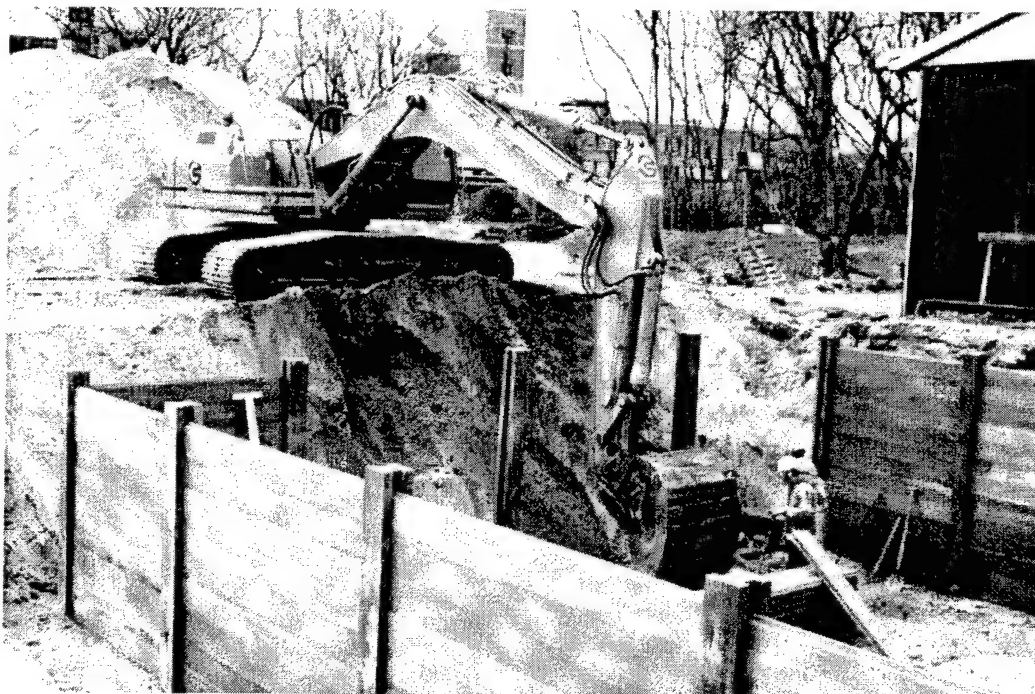
Figure 3.6   Last part of the digging phase. The wooden box is almost finished.
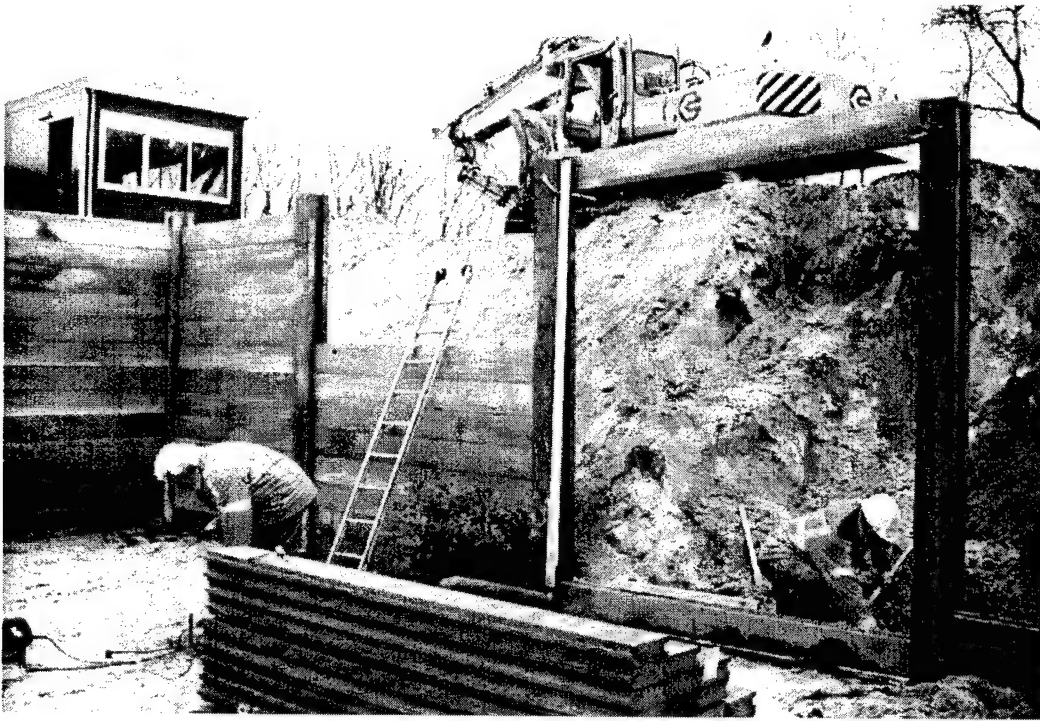
Figure 3.7   Closing of the wooden box.

Figure 3.8   The inside of the box is covered with a watertight plastic lining in order to prevent pollution from the outside (for example ground water) entering the box.
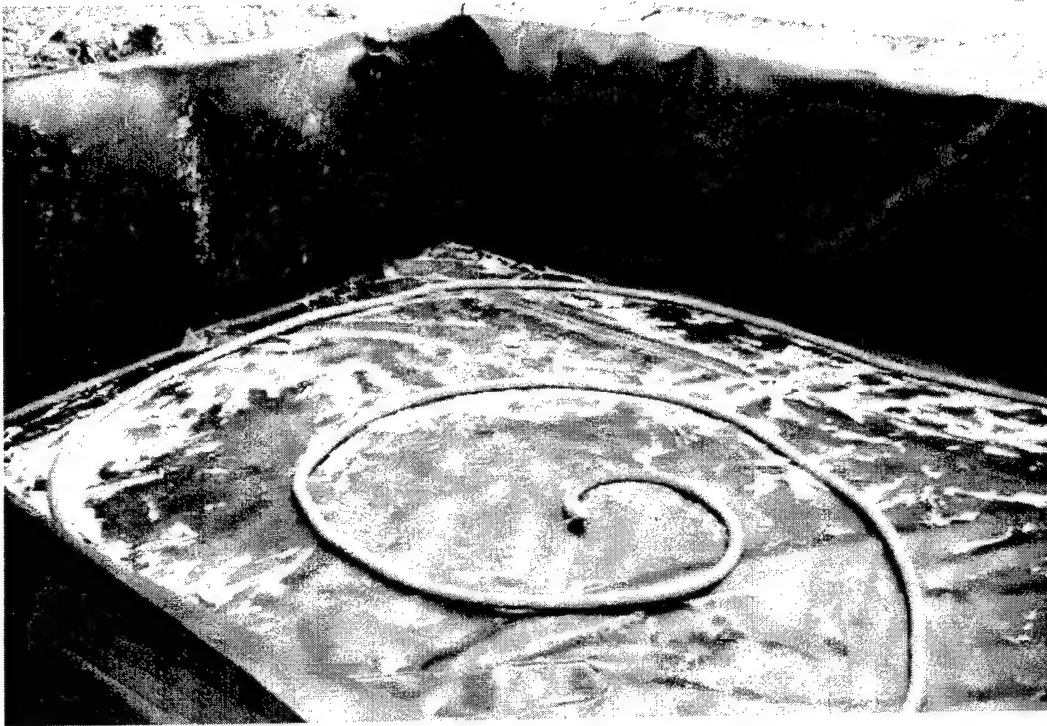
Figure 3.9   Installation of the drainage system in the wooden box.

Figure 3.10   Overview photo. The wooden box is now ready to be filled with sand.
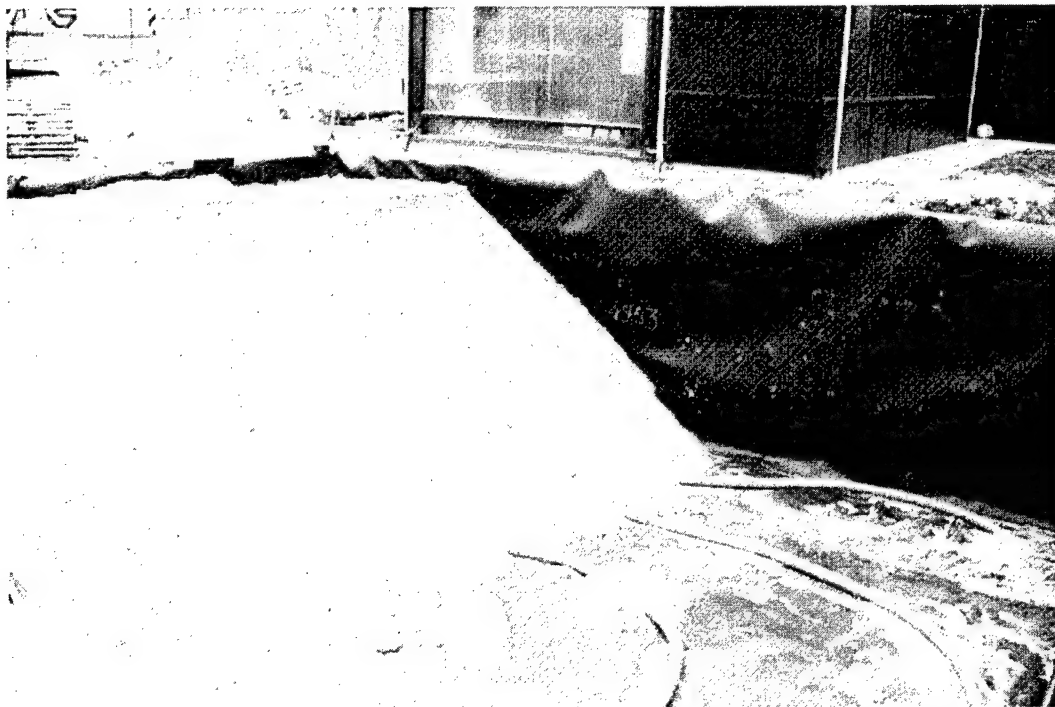
Figure 3.11   The wooden box is filled with clean (homogeneous) sand.

Figure 3.12   Installation of a square PVC tube running from the surface of one side to the bottom of the other side of the box, facilitating measurements of EM transmissions into the ground.

Figure 3.13   The wooden box is now ready to be used for ground-penetrating radar experiments. Presently the whole sandbox is covered by a huge tent.

Figure 3.14   The TNO-FEL impulse radiating antennas (IRA) constructed as part of the UWB-GPR systems.

4. HOM2000 TEST FACILITY

The other testfacility on the TNO test-site "Waalsdorp" is the HOM2000 test facility. The main purpose of this facility is the testing of commercial off the shelf (COTS) sensors that can be used for, and/or improved for mine-detection. Currently four sensor type are evaluated. Metal detectors, thermal infrared imaging detectors, and ground penetrating radars. For each sensor type, four systems are being tested

4.1 CONTRUCTION

The HOM2000 test facility consists of six test lanes. Each lane has a length of 10m, width of 3m and a depth of 1.5m. Every lane contains a different kind of soil representative of soil found in actual minefield locations. The kinds of soil used are, sand, clay, peat, soil with a high iron content, forrest soil, and rocky soil.

The test facility was constructed in such a way as to minimize external influences. No metal components are allowed inside a zone of 5m around the whole test-rig. Of course the test lanes themselves were also constructed without the use of any metal whatsoever. The facility was constructed along a north-south axis so that infra-red sensors would not have to take measurements in their own shadow. A groundwater level management system was also installed to be able to keep the groundwater table as a constant level. To facilitate accurate positioning of the sensors a non-metal measurement platform was designed and installed.

4.2 MEASUREMENT PLATFORM

The measurement platform is basically a big 10m long pvc/glassfibre bridge (tubular). Wheels on either end of the bridge are running on rails located at the head and foot of the test lanes. In this way the whole bridge can be moved over one of the test lanes. Once in position the bridge can be locked into place by a nylon bolt which fits in one of a series of holes every 0.05m. The sensor is mounted on a small belt driven trolley which is running on tracks over the full length of the bridge. The trolley has a small platform with mechanical adaptors to facilitate the mounting of the different sensors. At the end of the bridge, just outside the 5m metal-free zone, an electro-motor is connected to the belt pulling the trolley along the bridge. The maximum speed of the trolley is about

10m/minute. Braking is gentle but an emergency brake will stop the trolley abruptly in case it is needed.

To be able to record the position of the sensor acurately a laser distance meter is mounted on the far side of the bridge. The laser beam is reflected off a reflector mounted on the trolley and an accurate position reading and time-stamp is registered so many times per second. This way an accurate record of the position of the sensor is available for each sensor.

4.3 SURROGATE LAND-MINES

Inside every test-lane a number of surrogate land-mines and potential false alarms are buried at different locations and depths. These landmines are representative of what is typical in a real minefield. The emphasis however is mainly on anti-personel mines although some anti-tankmines are also burried (and of course a couple of false targets were also put in place). Control of the whole test-site and testlanes is quite strict as not to disturb the measurement conditions. Not only this but also security is an important reason to impose strict access control.

Once buried, the mines will not be dug up again. In fact a certain time (months!) has to be allowed to let the mines settle in the ground because it takes quite a while before the ground disturbance created when burrying the mines is deminished. To be able to do some exprimenting a small area of each testlane was set aside. This area can be used to burry sensors or reference objects. To be able to determine the character-sitics of the sensor with respect to "virgin" soil each test-lane also contains a small areas devoid of any surrogate landmines or false targets.

The construction of the surrogate landmines was done is such a was as to optimise it's resemblance to actual live mines. The resemblance to real mines is not just visual but care was also take to use materials whose electromagnetic and thermal properties closely match those of actual landmines. Details of the construction are not included here but could under certain conditions (constrains) be made available to others.

The following pictures show the construction of the measurement platform. In the picture you can see the antenna array and tick-wheel of a commercial GPR system which is mounted on the platform. For actual measurements the tick-wheel interface will be connected to the output of the laser position indicator.

Figure 4.1   A view of the HOM2000 measurement platform. In the background the tent covering the UWB-GPR test facility can be seen. In the distance one notices the TNO-FEL radar tower.
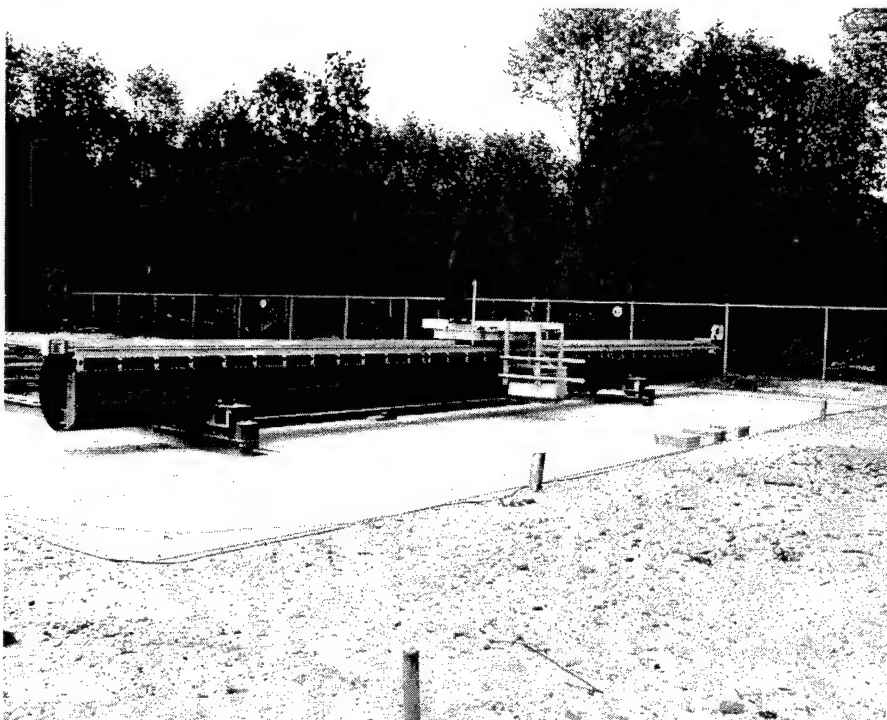

Figure 4.2   A view in the opposite (northerly) direction. The test lanes are just about visible. Due to the sensitive nature of the test facility more detailed pictures are not allowed.

Practical Application of Pattern Recognition Techniques

Hendrik Rothe
Arno von der Fecht
André Kasper
Thomas Rinder
e-mail: rothe@unibw-hamburg.de
Phone: 040/6541-2723 Fax: 040/6541-2743
Mess- und Informationstechnik, FB-MB, Universitaet der Bundeswehr
Holstenhofweg 85, D-22043 Hamburg

1 Introduction

This paper deals with three examples for pattern recognition and an estimation of the computational effort needed for the real time parts (predicition) of the following approaches:

- Discriminant Analysis

- Feedforward Neural Network

- *Sugeno* Fuzzy Classificator.

The first example is of the simple pattern matching type. A satellite photograph of our university from a freely distributed CD–ROM was used. We extracted the clock tower of the building where the chair for measurement and information technology is located and tried to find that place by correlation analysis. This pattern recognition works *without* features. We found the location of the clock tower with an accuracy of about ± 1 pixel (± 1 meter) in an photograph which actually displays 2×2 kilometers. The interesting point is here however, that one can use data freely available and a simple PC–platform with common numerical software like $MatLab^{TM}$.

In the second example we show our work in the field of in–situ inspection of high performance optics (e.g. degradation of airborne surveillance systems) and supersmooth surfaces, like wafers, hard discs and flat panel displays.
In this case the extraction of the correct feature variables and the appropriate classificators is of great importance for the performance of the pattern recognition system.
It is shown that empirical statistical moments and Discriminant Analysis perform best.

The third example of the paper is based on a range gated 3D laser radar system without scanner developed in our group. The design of the system is explained and measurement results are presented. This part of the paper explains, why one can use very simple features and straightforward classifiers, *if* the hardware of the pattern recognition system is specially designed *for* classification purposes. To our knowledge it is the first time that such an approach is published unclassified.

2 Pattern Matching in Satellite-photographs

2.1 List of Symbols

Symbol	Description
A, U	matrices
G	set
R_{xx}	auto correlation function
$E\ldots$	ensemble expectation operator
P_{xx}	*Fourier* transformed R_{xx}
$X(f)$	*Fourier* transformed x
$(\ldots)^*$	complex conjugated
$F\langle x \rangle$	*Fourier* transformed x
$F\langle x \rangle^{-1}$	inverse *Fourier* transformed x
R_{xy}	cross correlations function
P_{xy}	*Fourier* transformed R_{xy}
$\phi(f_k, f_l)$	coherency function
k, l, m_x, n_y	signed integer
m_x, n_y	signed integer
f_l, f_k	frequency

2.2 A simple Application of Pattern Matching

The detection of object positions in images is one of the basic problems in image processing. In this section the simple method of pattern matching will be discussed.

A rectangular limited region U will be searched for in an image A. A and U are matrices with the dimensions $r \times s$ and $m \times n$ such that $A \in G^m \times G^n$ and $U \in G^r \times G^s$ with $r, s, m, n \in N$. G is the quantity of gray values with $G = \{0, 1, 2, ..., 255\}$. The matrix U is nearly a copy of a submatrix V in A, the searched object. Finding the position of the submatrix V is an equivalent formulation of the problem. In preparation of the solution it will be assumed that A (the digitalized image) is a discrete piece of a normally distributed, place depending, 2- dimensional process (the real image). With this assumption the cross correlation function (CCF) can be used to detect the searched position. If two shifted images are similar, the CCF will be maximized at shift coordinates [1].

2.3 Mathematical formalism

The auto correlation function or sequence of a stationary process $x(n)$ is defined by

$$R_{xx}(m) = E\{x(n) \cdot x(n+m)\} \tag{1}$$

where $E\{...\}$ denotes the ensemble expectation operator and $x(n)$ is real valued. The power spectrum is formally defined as

$$P_{xx}(f) = E\{X(f) \cdot X^*(f)\} \tag{2}$$

where $X(f)$ is the *Fourier* transformation of $x(n)$ with

$$X(f) = \sum_{m=-\infty}^{+\infty} x(m) e^{-i2\pi fm} \langle x \rangle. \tag{3}$$

According to *Wiener Chintchine* theorem it is the *Fourier* transform (FT) of the auto correlation function and may consequently be written as

$$P_{xx}(f) = \sum_{m=-\infty}^{+\infty} R_{xx}(m) e^{-i2\pi fm} = F\langle R_{xx}\rangle. \tag{4}$$

Based on the basic definitions (1) and (4) the cross correlation function is defined as

$$R_{xy}(m) = E\{x(n) \cdot y(n+m)\} \tag{5}$$

and the cross power spectrum as

$$P_{xy}(f) = \sum_{m=-\infty}^{+\infty} R_{xy}(m) e^{-i2\pi fm} = F\langle R_{xy}\rangle \tag{6}$$

if $x(m, n)$ and $y(m, n)$ are real valued. The equations (1) to (6) apply for one – dimensional processes.

The generalization for two – dimensional functions $x(m, n)$ and $y(m, n)$ is simply given by

$$R_{xy}(k, l) = E\{x(n, m) \cdot y(n+k, m+l)\} \tag{7}$$

$$P_{xy}(f_k, f_l) = E\{X(f_k, f_l) \cdot Y^*(f_k, f_l)\} \tag{8}$$

Finally the coherency function $\phi(f_k, f_l)$ is defined as

$$\phi^2 = \frac{|P_{xy}(f_k, f_l)|^2}{P_{xx}(f_k, f_l) \cdot P_{yy}(f_k, f_l)} \tag{9}$$

Equation (9) contains mainly information about the similarity of two two – dimensional functions or sequences. For instance, the coherency function of two two – dimensional, identical but shifted functions $h(m, n)$ and $g(m, n)$ with

$$F\langle h\rangle = F\langle g\rangle e^{-i2\pi(f_k \cdot m_x + f_l \cdot n_y)} \tag{10}$$

is constant with the value one and leads to the identity

$$P_{gh} = P_{gg} e^{i2\pi(f_k \cdot m_x + f_l \cdot n_y)}. \tag{11}$$

Otherwise the shift m_y, n_y can be calculated from the phase term. In the other case the coherency function is nearly zero. This means that there is a random relative phase distribution between the two functions [2]. If the functions $h(m, n)$ and $g(m, n)$ are stochastically distributed, without periodical components and mean value, the function $R_{gg}(m, n)$ is globally maximized at the point $(0, 0)$. So the calculation of the shift m_y, n_y can be done simply with the inverse *Fourier* transformation of equation (11)

$$R_{gh}(k, l) = R_{gg}(k + m_x, l + n_y) \tag{12}$$

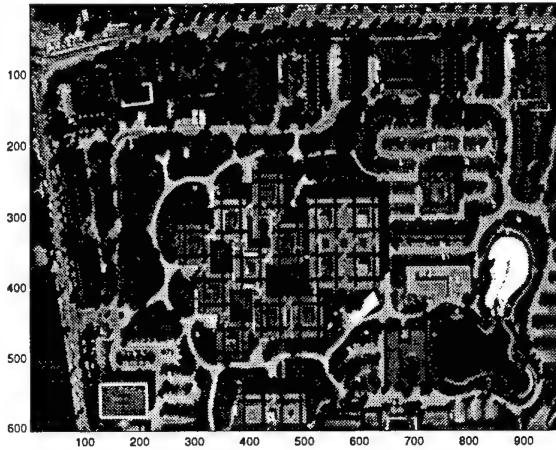and with a maximum search at $R_{gh}(k, l)$ [3].
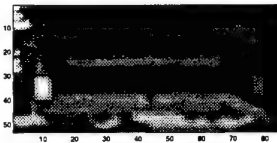
Figure 1: Satellite photography

Figure 2: Satellite photography - searched region

2.4 A simple application

With the assumptions given in section 2.2 a simple application will be shown. The image A is shown in Figure 1 and the searched region U in Figure 2.

Figures 1 and 2 are noncontinous, equidistant sampled regions. In section 2.3 it has been shown that there are two ways to calculate function $R_{AU}(k, l)$, see equation (7) and the inverse of equation (8). Equation (7) is simple to evaluate on A and U. It is a non causal convolution. With the assumption that the regions are periodically continued in both dimensions, with a loss of accuracy (ignoring bounding effects) the inverse of equation (8) can be used. To fit the smaller region U in dimensions of A the simple zero padding methode can be used, with U' as the zero padded region U

$$U' = \begin{pmatrix} u_{1,1} & . & . & u_{1,s} & 0 & . & 0 \\ . & . & . & . & 0 & . & 0 \\ . & . & . & . & 0 & . & 0 \\ . & . & . & . & 0 & . & 0 \\ u_{r,1} & . & . & u_{r,s} & 0 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & . & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (13)$$

with $U' \in G^m \times G^n$. So the final equation can be written as

$$R_{AU}(k, l) = F^{-1} \left\langle F \left\langle A - \overline{A} \right\rangle \cdot F \left\langle U' - \overline{U} \right\rangle \right\rangle . (14)$$

With the equation (13) the shift coordinates between the regions U and V are $k = m_x$ and $l = n_y$ which maximize $R_{AU}(k, l)$.
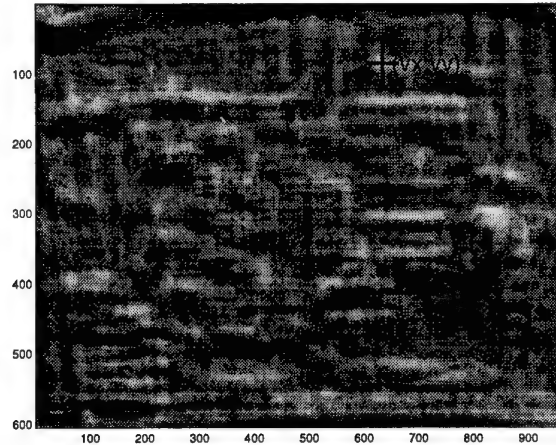
Figure 3: CCF

2.5 Application hints

Equation (14) is easy to compute using the $Matlab^{TM}$ computation system. The following script ($Matlab^{TM}$ Version 5.2, image processing toolbox) shows the computation of Figure 3.

```
[b1 map1] = imread('campgray.pcx');
[b2 map2] = imread('campgraypat.pcx');
A = double(b1);
U = double(b2);
[r s] = size(b2);
B = zeros(m,n);
B(1:r,1:s)= U;
F1 = fft2(A - mean(mean(A)));
F2 = fft2(B - mean(mean(B)));
C = real(ifft2(F1 .* conj(F2)));
[vx vy] = find(max(max(C)) == C);
```

The variable C represents $R_{AU}(k, l)$ and is theoretically real valued but numerical errors of the computation system produce small imaginary values in C. The computation speed can be increased, if m, n, r, s have the logarithmical base 2. So the computation of the Discrete *Fourier* Transform (DFT) can be changed to Fast *Fourier* Transform (FFT) [3].

The most accurate, but slower way to calculate $R_{AU}(k,l)$ is equation (7) which is shown in the next $Matlab^{TM}$ script.

```
[b1 map1] = imread('campgray.pcx');
[b2 map2] = imread('campgraypat.pcx');
A = double(b1) - mean(mean(b1));
U = double(b2) - mean(mean(b2));
C = conv2(A,flipud(fliplr(U)));
[vx vy] = find(max(max(C)) == C);
```

3 Classification of Surfaces by using Backscattering

Measuring the distribution of light scattering back from a illuminated surface is a fast and non-contact approach to characterize surfaces [4], i.e. measuring surface roughness, detection of defects and contaminations, characterization of surface textures. The bidirectional reflectance distribution function (BRDF) can be used to describe the scattering of light from a illuminated surface. The BRDF is defined as the ratio of scattered radiance L_s at an particular solid angle Ω_s to the surface irradiance E_i.

$$BRDF = \frac{L_s}{E_i} \approx \frac{\Phi_s}{\Phi_i cos\theta_s \Omega_s} \qquad (15)$$

As illustrated in Figure 4 the incident ray is defined by its incident angle θ_i and the scattered ray is defined by the azimuth angle ϕ_s (measured from plane of incidence), the scatter angle θ_s and the solid angle Ω_s.
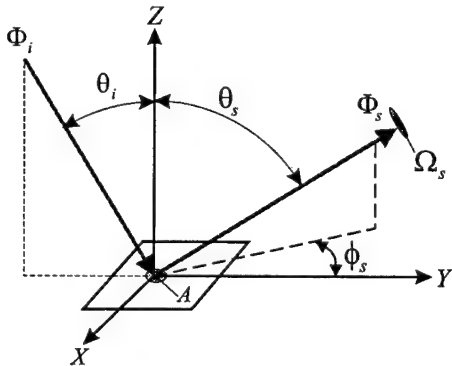
Figure 4: Definition of BRDF.

Apart from a few limited scatter theories, pattern recognition methods must be employed to link the BRDF to surface properties.

3.1 Sample data

The angle resolved BRDF data used in the following sections were taken from 11 surfaces, with different microtopographies. Each surface represents one group. A group consists of two data sets with 50 observations each. One data set (recognition set) is necessary for the learning part of the classification and the other set (prediction set) is intended for testing the classification. In Figure 5 the 11 group means are shown. Only the scatter distribution in the plane of incidence ($\phi_s = 0$) is evaluated. The angle of incidence was fixed to 38°.
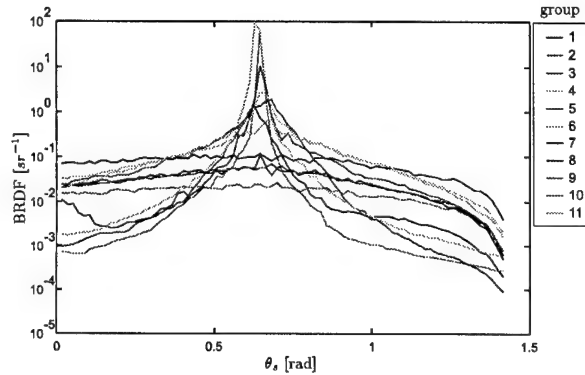

Figure 5: Plot of group means.

3.2 Statistical moments

The advantage of using statistical moments as feature variables ist that they are easy and fast to calculate. A drawback of statistical moments is that they only describe the statistical distribution of the BRDF independently of the function parameters θ_s, ϕ_s.
The computed statistical moments of the recognition set are shown in Figure 6.
The statistical moments of group 8, 9 and 10 do not show any significant difference, therefore classification errors are very likely for these groups.

3.3 Circular moments

To overcome the mentioned drawback of linear statistical moments, modern circular statistics can be employed to define circular trigonometric
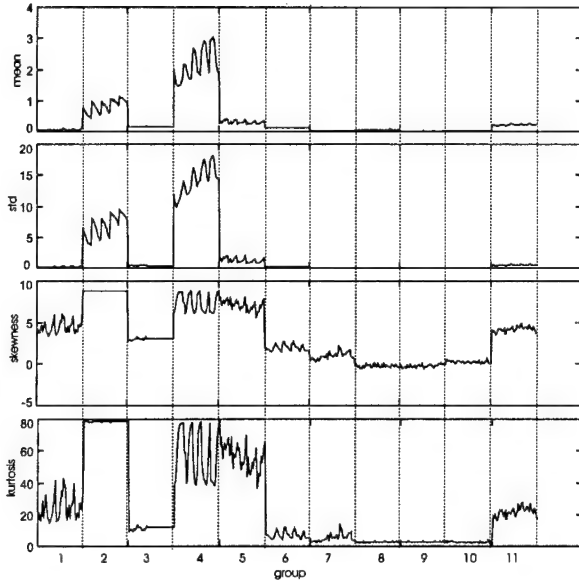
Figure 6: Plot of linear statistical moments vs. observations.

moments. Circular statistic is based on vectorial or axial data on the unit circle [5]. Therefore, in order to employ circular statistics to BRDF data, it is necessary to project the intensity values of the BRDF on the unit circle. This can be done by interpreting the intensity value of a certain solid angle as a summmation of elementary events, e.g. number of photons reaching the detector. With this interpretation one can consider BRDF data as a circular distribution, as shown in Figure 7.
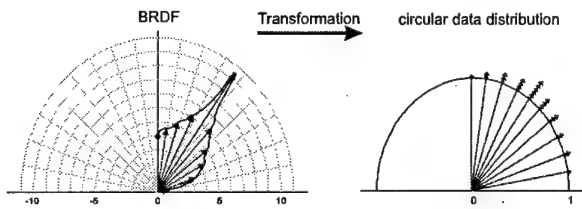
Figure 7: Interpretation of BRDF as circular data distribution, the triangles in the right figure symbolize the number of elementary events.

Mathematically, the transformation is given by the angular frequency value n_j,

$$n_j = \frac{BRDF(\theta_j) - min(BRDF)}{max(BRDF) - min(BRDF)} * N_\Phi$$

$$n_s = \sum_{j=1}^{N_{\theta_s}} n_j$$

$$(16)$$

where θ_j is the j-th scattering angle θ_s ($j = 1..N_{\theta_s}$), min(BRDF) and max(BRDF) are the minimum and maximum values of the BRDF data, respectivly, N_Φ is the number of elementary events the intensity value of the BRDF is divided into, n_s is the number of elementary events in the sample data.

Now the sample *mean direction* $\bar{\theta}$ can be calculated

$$C = \sum_{j=1}^{N_{\theta_s}} n_j \cos \theta_j \quad S = \sum_{j=1}^{N_{\theta_s}} n_j \sin \theta_j$$

$$R^2 = C^2 + S^2$$

$$(17)$$

$$\bar{\theta} = \begin{cases} tan^{-1}(S/C) & S > 0, \bar{C} > 0 \\ tan^{-1}(S/C) + \pi & C < 0 \\ tan^{-1}(S/C) + 2\pi & S < 0, \bar{C} > 0 \end{cases}$$

which should be equal to the angle of incidence θ_i for a specular BRDF. By defining the *mean resultant length* $\bar{R} \in (0,1)$ the *sample circular variance* $V \in (0,1)$ and the *sample circular standard deviation* v can be calculated.

$$\bar{R} = R/n_s$$

$$(18)$$

$$V = 1 - \bar{R} \quad v = sqrt{-2 * log(1-V)}$$

From the *centered sample trigonometric moments* m_p

$$m_p = 1/n_s * \sum_{j=1}^{N_{\theta_s}} n_j \cos p * (\theta_j - \bar{\theta}) \qquad (19)$$

the *sample circular dispersion* σ (a measure of spread)

$$\sigma = (1 - m_2)/2 * \bar{R}^2 \qquad (20)$$

and measures of *skewness s* and *kurtosis k* can be defined [5].

$$s = m_2 * sin(\mu_2 - 2\bar{\theta})/(1 - \bar{R})^{\frac{3}{2}}$$
$$k = (m_2 * cos(\mu_2 - 2\bar{\theta}) - \bar{R}^4)/(1 - \bar{R})^2 \qquad (21)$$

The value μ_2 is calculated from the second *uncentered sample trigonometric moment*

$$\bar{C}_p = 1/n_s * \sum_{j=1}^{N_{\theta_s}} n_j \cos(p\theta_j)$$

$$\bar{S}_p = 1/n_s * \sum_{j=1}^{N_{\theta_s}} n_j \sin(p\theta_j)$$

$$(22)$$

$$\mu_2 = \begin{cases} tan^{-1}(\bar{S}_2/\bar{C}_2) & \bar{S}_2 > 0, \bar{C}_2 > 0 \\ tan^{-1}(\bar{S}_2/\bar{C}_2) + \pi & \bar{C}_2 < 0 \\ tan^{-1}(\bar{S}_2/\bar{C}_2) + 2\pi & \bar{S}_2 < 0, \bar{C}_2 > 0 \end{cases}$$

The computed circular moments of the recognition set are shown in Figure 8.

As one can see from the feature plot the circular moments of group 8 and 9 do not show any significant difference, therefore classification errors are very likely for these groups.
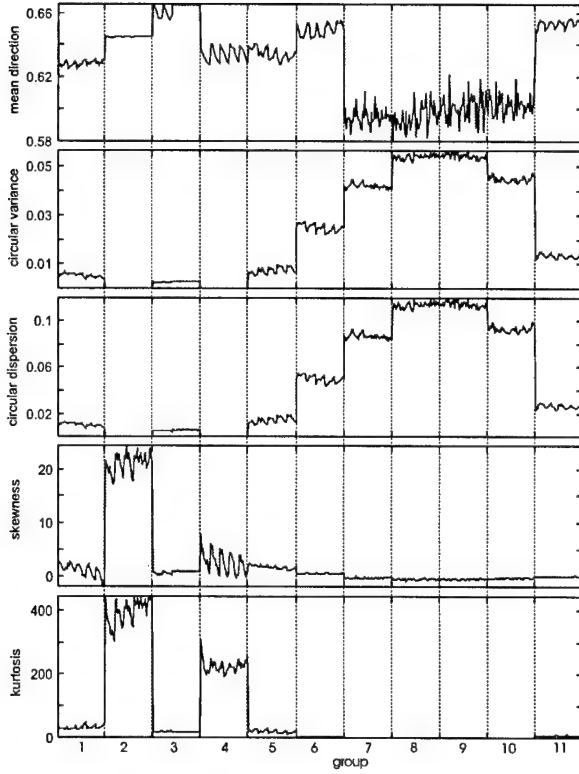
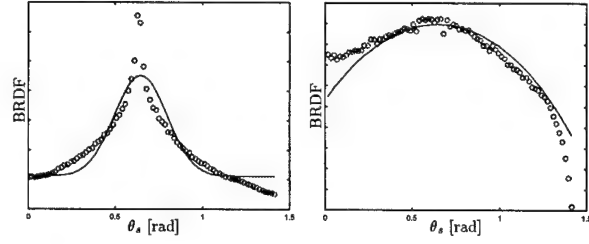Figure 8: Plot of circular moments vs. observations.

Figure 9: Plot of two approximations with *Phong* model, l. of a specular BRDF, r. of a smooth BRDF.
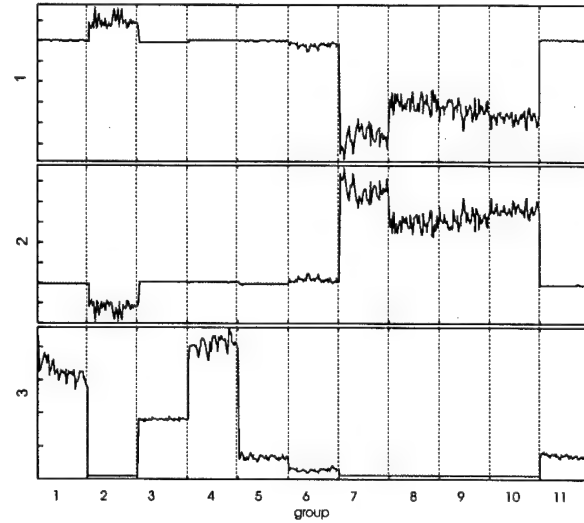
Figure 10: Plot of coefficients of *Phong* model vs. observations.

3.4 Features based on BRDF models

Another approch to feature extraction of BRDF data is based on the approximation of BRDF models and using the estimated parameters as feature variables. Quite a lot of BRDF models were suggested by the computer graphics community. The definition of two models will be given. One of the earlier and still quite popular models ist the *Phong* model [6]. A slightly modified model was introduced by [7] and is defined as a sum of a diffuse part and a specular part.

$$f_1(\theta_s, \theta_i) = a_1 \frac{1}{\pi} + a_2 \frac{a_3 + 2}{2\pi} cos(\theta_s - \theta_i)^{a_3} \quad (23)$$

The diffuse part depends on a_1 and the specular part depends on a_2 and a_3. Two examples for the *Phong* approximation of a diffuse and a specular BRDF are shown in Figure 9.

The least squares approximation was done with the Quasi-*Newton* algorithm. The computed feature variables (a_1, a_2, a_3) of the recognition set are shown in Figure 10.

The model of *Meister* [8] allows a more accurate description of the diffuse part and the specular part. It is defined as a summation of a constant diffuse part a angle dependent diffuse part and a specular part.

$$f_2(\theta_s, \theta_i) = a_0 +$$

$$a_1(\theta_i^2 + \theta_s^2) + a_2(\theta_i \theta_s)^2 + a_3 \theta_i \theta_s cos(\phi_s) + \quad (24)$$

$$a_4 e^{a_5(\theta_i \theta_s)^2} e^{-a_6(\theta_s - \theta_i)^2}$$

The coefficient a_0 gives the constant diffuse part of the sample. The angle dependent diffuse part is modeled by the coefficients a_1, a_2, a_3 and the coefficients a_4, a_5 describe the amplitude and a_6 stands for the width of the specular part. According to the model the coefficients a_0, a_4, a_5 and a_6 should be positive. Figure 11 shows two aproximations of the *Meister* model.

The least squares approximation was done with the Quasi-*Newton* algorithm. The computed feature variables $(a_1 - a_6)$ of the recognition set are shown in Figure 12.
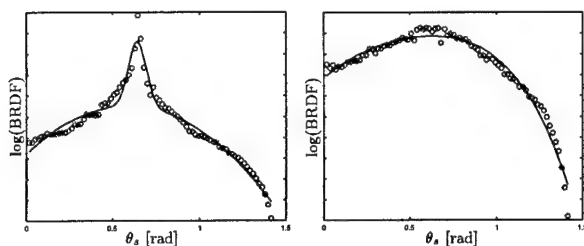
Figure 11: Plot of two approximations with *Meister* model, l. of a specular BRDF, r. of a smooth BRDF.
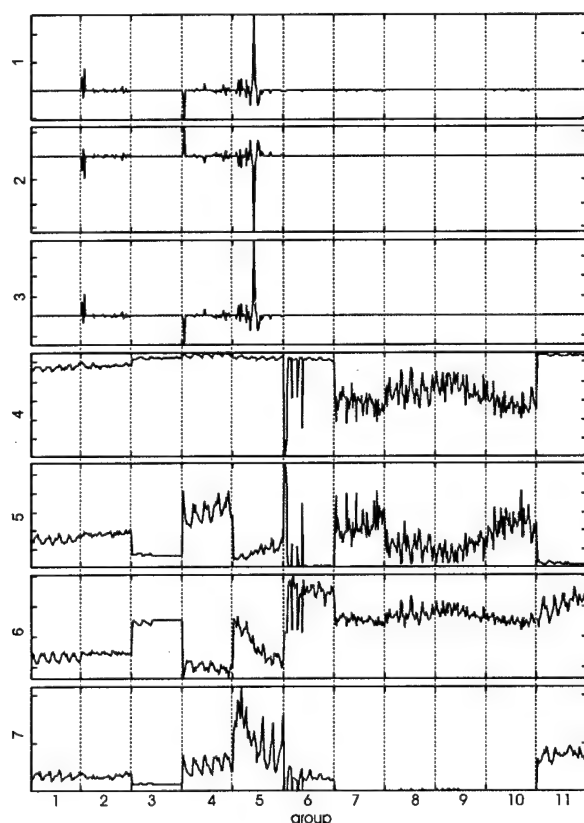

Figure 12: Plot of coefficients of *Meister* model vs. observations.

3.5 · Features based on spline approximation

In the last section the BRDF was approximated by a certain function model. In the case of not knowing the shape of the approximation function, spline - curves can be employed to approximate the light distribution. For this example the B-spline form is employed [9]. The basis functions of a B-spline curve are called B-spline functions. A B-spline function of order k with the ordered knot sequence T (the parameter values of the B-spline function)

$$T = (t_0, t_1, t_2, ..., t_{n+k-1}, t_{n+k}) \tag{25}$$
$$t_1 \leq t_2 \leq ... \leq t_{n+k}$$

can be defined as recurrence relation, as follows:

if $k = 1$

$$N_{i,1}(t) = \begin{cases} 1 \text{ if } t_i \leq t \leq t_{i+1} \\ 0 \text{ else} \end{cases}$$

if $k > 1$ $\qquad\qquad\qquad\qquad (26)$

$$N_{i,k}(t) = \frac{t - t_i}{t_{i+k-1} - t_i} N_{i,k-1}(t) +$$

$$\frac{t_{i+k} - t}{t_{i+k} - t_{i+1}} N_{i+1,k-1}(t).$$

A B-spline curve of order k is given as linear combination of spline functions $N_{i,k}$.

$$B_k(t) = \sum_{i=1}^{n} a_i N_{i,k}(t) \tag{27}$$

The coefficients a_i are called the *control points* or *de Boor points* for the curve. The smoothness of the B-spline curve can be controlled by the knot sequence T, i.e. for a jump in the first derivative at t_3 one would repeat the knot value t_3 $k - 1$ times in the knot sequence. This B-spline property can be used to improve the approximation of BRDFs with a strong specular part.

The approximation of a given set of points P_i (i.e. a measured BRDF) can be expressed as the minimization of the sum of the square errors D_i.

$$D_i = B_k(t_i) - P_i \quad D^2 = \sum_{i=0}^{n} D_i^2 \rightarrow \min \tag{28}$$

Partially deriving D^2 to a_i leads to a linear system of equations. Solving this system of equations gives the desired control points a_i. An example of two B-spline approximations of differently shaped light distributions are shown in Figure 13.

The control points a_i of the approximated B-spline curve can be used as feature variables when the knot sequence T and the order k are constant for the data set. The computed feature variables of the lerning set based on a B-spline curve of order four with six intervals and one possible jump in the first derivative at $\theta_s = \theta_i$ are shown in Figure 14.
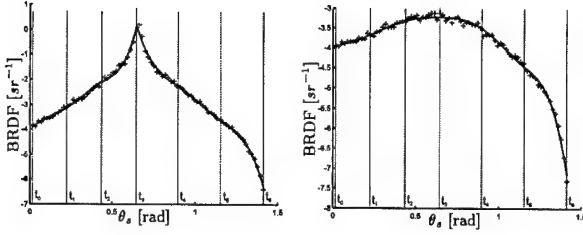
Figure 13: Two approximations with fourth order B-splines and six intervals, l. with jump in the first derivative at t_3, r. with continous derivates.
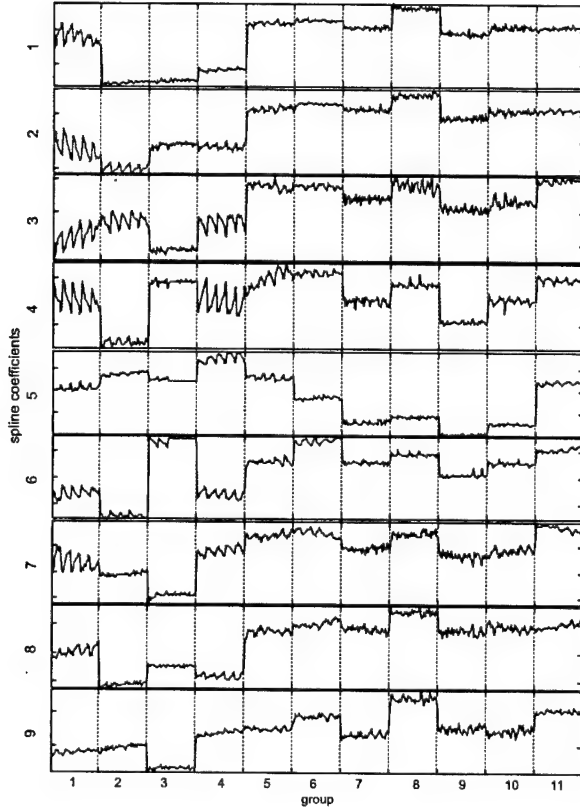

Figure 14: Plot of coefficients of B-spline approximation vs. observations.

3.6 Classification results

Now the ability of the defined BRDF features to significant characterize a class of BRDFs is to be investigated. A first impression of the quality of the approximation based features can be obtained by looking at the approximation errors. By expressing the approximation error as follows

$$TS_{\chi^2} = \sum_{i=1}^{n} \frac{(P_i - f(t_i))^2}{\sigma_i^2}, \qquad (29)$$

the statistical χ^2 test can be used to check the quality of the fit. If the value TS_{χ^2} is larger than the corresponding quantile value the fit should be considered non significant. The degree of freedom for the χ^2 quantile is given by the number of approximated data points minus the number of free parameters. Figure 15 shows the TS_{χ^2} values for the approximations applied to the recognition set. The straight lines symbolize the corresponding χ^2 quantile values. From the plot of Figure 15 it can be stated that the B-spline approximation gives the best result according to the approximation errors. Apart from the observations of group two and four the TS_{χ^2} values of the B-spline and the *Meister* approximations are below the quantile values. Therefore they can be considered as significant due to the χ^2 test. However, the real discrimination performance of the features can only be qualified by evaluating the classification results. Table 1 shows the summarized classification results for different classification methods. For each algorithm the classification results of the recognition set and the test set are given. The first three algorithms are based on neural networks with the following start configurations:

- The perceptron net was initialized with 11 neurons in the output layer.

- The feedforward net was initialized as two layer configuration with the number of neurons in the input layer equal to the number of feature variables and 11 neurons in the output layer. The position of the maximum activation in the output layer is considered as classification result.

- The radial basis function network was initialized in the same manner as the feedforward network.
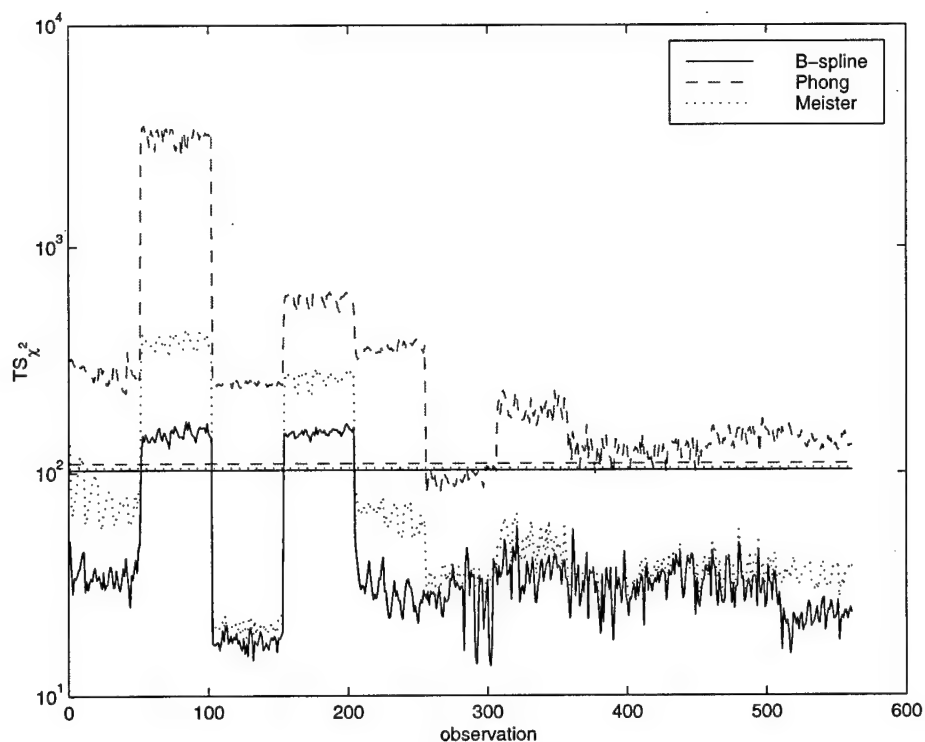
Figure 15: Plot of TS_{χ^2} approximation errors vs. observations.

Table 1: Table of classification results, shows the correct classifications in percent.

classification algorithm	empirical moments	circular moments	*Phong* model coefficients	*Meister* model coefficients	B-spline coefficients	all features
Perceptron						
recognition set	75.3%	30.0%	22.2%	52.2%	90.2%	80.4%
prediction set	73.3%	26.7%	22.5%	39.6%	82.9%	62.4%
Feed Forward						
recognition set	92.4%	82.9%	62.5%	78.9%	94.7%	93.3%
prediction set	95.0%	86.5%	73.8%	93.6%	98.6%	99.3%
Radial Basis Function						
recognition set	100.0%	98.4%	100.0%	100.0%	100.0%	100.0%
prediction set	82.2%	15.9%	51.5%	10.3%	44.4%	10.3%
Discra						
recognition set	98.4%	86.6%	90.4%	81.8%	99.5%	99.8%
prediction set	99.1%	81.8%	88.4%	74.9%	97.7%	99.8%

The following conclusions can be drawn from the classification results:

- The discriminant analysis and the feedforward neural network work best with the introduced features, whereas the discriminant analysis is slightly better, especially for worse features.

- The classification performance of the perceptron neural network is not sufficent.

- The feedforward neural network gives a very good classification result for the learning step but for the test set the classification performance is very poor. This might be improved by changing the network configuration, i.e. by including additional hidden layers.

- The empirical statistical moments and the B-spline coefficients give the best classification results independent of the classification algorithm used.

- The *Phong* and *Meister* approximation based features do not show the desired discriminatory power. A possible cause is that the nonlinear optimization might lead to different parameter configurations for almost indentical BRDFs. This could be improved by applying better start parameters and by checking the parameter configuration during the optimization process for validility due to the BRDF model.

4 Recognition of Objects in 3D Laser Radar Images

4.1 Basics of Laser Range Gating

There are different possibilities to generate 3D-images at a long range (> 100m) with a pulse laser:

- scanning camera system that combines a scanning system (rotating mirror) with a single detector (Avalanche Photodiode APD) [10],

- APD array receiver [11, 12],

- gated viewing camera [13, 14].

These systems could be used in automatic target recognition, aimpoint selection, target tracking and obstacle avoidance [15, 16, 17, 18].

What we will mainly examine here are imaging laser radar systems, which are based on a gated viewing camera. Figure 16 demonstrates the principle of laser range gating. The scene is illuminated with a pulsed laser and synchronized to a gated viewing camera. The delay of camera exposure and the duration of the exposure can be controlled with an accuracy of 1ns. It is possible to achieve exposure times Δt_{exp} of 3ns. With

$$\Delta R = \frac{c * (\Delta t_{exp} + \Delta t_{puls})}{2}, \qquad (30)$$

where c is the velocity of light and Δt_{puls} is the pulse width of the laser, it can be recognized that the camera frame includes only the scene at a given range with a depth ΔR of 1.5m with a pulse width Δt_{puls} of 7ns. The very fast shutter speed of 3ns can be achieved with a MCP (Multi Channel Plate), Generation II. The frame is generated only from the reflected photons from the target slice T5 (Figure 16). The photons of the scattered light outside the target slices T1-T4,T6-T9 (Figure 16) do not reach the camera inside the exposure time and will therefore not be added to the image.

The timing can be controlled at every laser shot and consequently one gets illuminated slices at different ranges with different depths. The combination of all slices gives a 3D-image. But normally it is not useful to combine the range gated images because an image processing system can extract faster more information out of the single range gated images.

4.2 Eye-safety

The laser radar system mentioned in section 4.1 has one big disadvantage. It is operating in the visible region at a wavelength of 532nm. For an "all weather" long range imaging laser radar with the requirement of high laser power it is useless due to the extreme eye-hazard near the maximum sensitivity of the human eye. These laser radars working with a MCP are nowadays only practical for underwater applications like mine detection or controlling of autonomous underwater robots.

Achieving more laser power without an eye-hazard is only possible beyond a wavelength 1.4μm. In this spectral region there is hardly any transmission between the cornea and the retina, so that the incoming light is not focussed on the

retina. The laser power limits beyond $1.4\mu m$ are nearly the same for the eye and the skin. Providing a class 1 laser radar system means not to exceed the AEL (Accessible Emission Limit) outside the transmitter optic. A class 1 system can operate without any safety rules [19].

For laser radar systems the AEL depends on the number of pulses N that could reach the cornea inside a time period of 100s:

$$AEL_{multipulse} = N^{-\frac{1}{4}} AEL_{singlepulse}. \qquad (31)$$

Table 2 shows that the permitted laser power in the eye-safe spectral region beyond $1.4\mu m$ can be much higher than in the visible.

For an "all weather" long range imaging laser radar system it is absolutely necessary that it operates in the eye-safe region. High power laser sources like Raman shifted Nd-YAG-, Nd-YAG OPO(Optical Parametric Oscillator)- or Erbium doped fiber lasers are available on the market, but imaging receiver working beyond $1.1\mu m$ are very rare.

4.3 Design of an eye-safe, imaging laser radar

The development of an eye-safe, imaging, scannerless laser radar based on gated viewing requires another shutter principle because the use of a MCP in the eye-safe region of $1.5\mu m$ [20] seems not realistic in the near future [21]. The requirements for such a system are:

- high power eye-safe laser source for single shot illumination,

- high sensitive detector at a wavelength of $1.5\mu m$,

- very fast shutter in the nanosecond region,

- precise delay generator with a stability in the picosecond region.

The whole system design is shown in Figure 17. We built a complete computer controllable system including a BigSky Nd:YAG laser CFR 400 OPO, an InGaAs camera SU128-1.7RT CCIR from Sensor Unlimited [22], a delay generator EG&G model 9650A and an electro-optical modulator EOM from Conoptics.

The InGaAs camera has a quantum efficiency of 68% at a wavelength of $1.5\mu m$ and a detectivity D^* of $10^{13} \frac{cm\sqrt{Hz}}{W}$. The camera works at a video rate of 25Hz and triggers the laser. At this rate the laser achieves a pulse energy of 60mJ and a pulse width of 7ns. The pump energy of the flashlamp pumped laser can be controlled via RS232, so that it is possible to increase or decrease the laser output energy every laser shot. The Q-switch sync output is used to get the time of laser start and to trigger the delay generator which gets the delay time and pulse duration via RS232. The electrical output impulse with a given delay and pulse width opens the EOM shutter for a duration between 18ns and 500ns corresponding to the input pulse width. Opening the EOM shutter means that the dc voltage level at the variable dc voltage supply provides the half wave voltage. The EOM shutter we use here is of the transverse field type with the material lithium tantalate. For the EOM shutter we have calculated and designed a special objective. It includes an analyzer, a polarizer and an IR bandpass filter with the FWHM optical bandwidth of 30nm. The focal length of the objective is 400mm.

The whole system design integrates an eye-safe rangefinder working with an InGaAs APD as detector, which communicates with the control computer via centronics. The rangefinder does not work with last pulse technique, but we can set a range gate of any size and manage to get the range also at bad weather conditions as heavy rain or fog. The laser radar control software could use the range information from a target to set the corresponding delay of the delay generator. The observation of the target per range gating system could start at this location.

We have finished the system design yet, but it was not possible to provide outdoor experimental results. Therefore we use for the example of object recognition and classification section 4.5 a range gating sequence provided with a gated viewing system including a MCP.

4.4 Long range experiments with a gated viewing system at the wavelength of 532nm

The target used for the long range experiments was a truck (Figure 18(a)) with a special wire rack (Figure 18(b)). These two images were taken with a video camera Sony CCD-TRV101E. The truck painted with olive green drab is standing in a distance of 740m to the measurement house. It is nearly a Lambertian target with a reflectivity $\rho = 0.1$. The wire rack consists of a wooden frame

including following crossbars from top to buttom:

type	diameter [mm]
copper wire	3.3
cable	6
cable	13
aluminum pipe	20
aluminum pipe	30
aluminum pipe	60

Figure 19 shows a sequence of range gated images. Every image in the sequence is a mean value image of 20 single range gated images with the same delay and duration of the gate. The whole sequence consists of 8 range slices corresponding to the 8 range values (grey scale bar in the range image Figure 20(b)). Because the images were taken with 25Hz it took 6.4s to get the sequence and the resulting range image Figure 20(b).

Only 6 of the whole 8 images presented in Figure 19 because the range slices starting at 0.58km and 1.47km range do not contain information. The duration of the gate was 993ns for every image slice that means with (30) a range depth ΔR of 150m. Table 3 displays the corresponding system parameters for the image sequence of Figure 19. There are some discernible peculiarities in the image sequence. In the first image of the sequence the plants are blurred, because the focus is set to the truck in the range of 740m. Consequently the visible shadow of the plants in the second image is sharp. The solution of this problem is the integration of a focus control or image processing like an image restauration with a Wiener filter.

The next peculiarity can be seen in Figure 19(e). The image has a lot of overlight in the left top side which causes scattered light in the objective. The reason for this is a retroreflector in the middle of the target. But the existence of such a hot spot illustrates the effective of range gating. The range slice of the truck is not influenced by the hot spot. Consequently, the truck can be detected clearly.

The different range slices are combined to a range image. The range image Figure 20(b) is calculated with a simple threshold of 20. Every grey value is corresponding to a range as it can be seen in the grey scale bar in the right side of the image. The grey value of 0 (black) at a location (pixel (x,y)) is used if there is no grey value in the sequence higher than the threshold. The choice of a threshold depends on the maximum noise value after building the mean value image.

For the computer controllable system it is important to consider the enormous influence of the choosen MCP gain on the image noise. A high gain provides a very strong excess noise as it can be seen in Figure 20(a). This is a corresponding single image to the mean value image (20 single images) of Figure 19(b). The system works here with maximum MCP gain of 18000.

4.5 Example of Object Recognition with a Laser Radar Range Gated Sequence

4.5.1 Introduction

In the example it will be shown that object recognition in a laser radar image sequence has a lot of advantages in comparison to normal grey value images. That means that in a lot of cases an automatic object detection, the classification of these objects and the assignment to different classes is only possible while using range gating.

The nature of range gating is that only a special object is illuminated and the background and foreground are both black. For digital image processing that means that the first step of an object recognition algorithm, object segmentation, can be done by simple thresholding. While looking at the image sequence in Figure 19 it can be imagined that the normal grey value image of this scene does not allow to detect the truck because the grey values of the bush in the foreground and the forest in the background include nearly the same scope.

The next big advantage is the knowledge of the distance of a range slice. With the distance and the main parameter of the objective and the camera it is no problem to calculate the real size of an object. The same target in different ranges has always a different size in pixel in the image. In the range gated images the correct size in meter of the target can be calculated and therefore for an object recognition algorithm the object size is range independent.

In the next subsection we want to discuss the necessary steps of an object recognition algorithm for the automatic detection of the truck in the range gating sequence of section 4.4 in detail. It can be imagined that the same algorithm is also capable to distinguish different types of trucks, tanks, helicopters or different weapon systems. The object features we use here are only a small choice of the possible features. Especially for the classification of geometric similar objects the signature moments could be a promising choice. In the example we do not use the object signature.
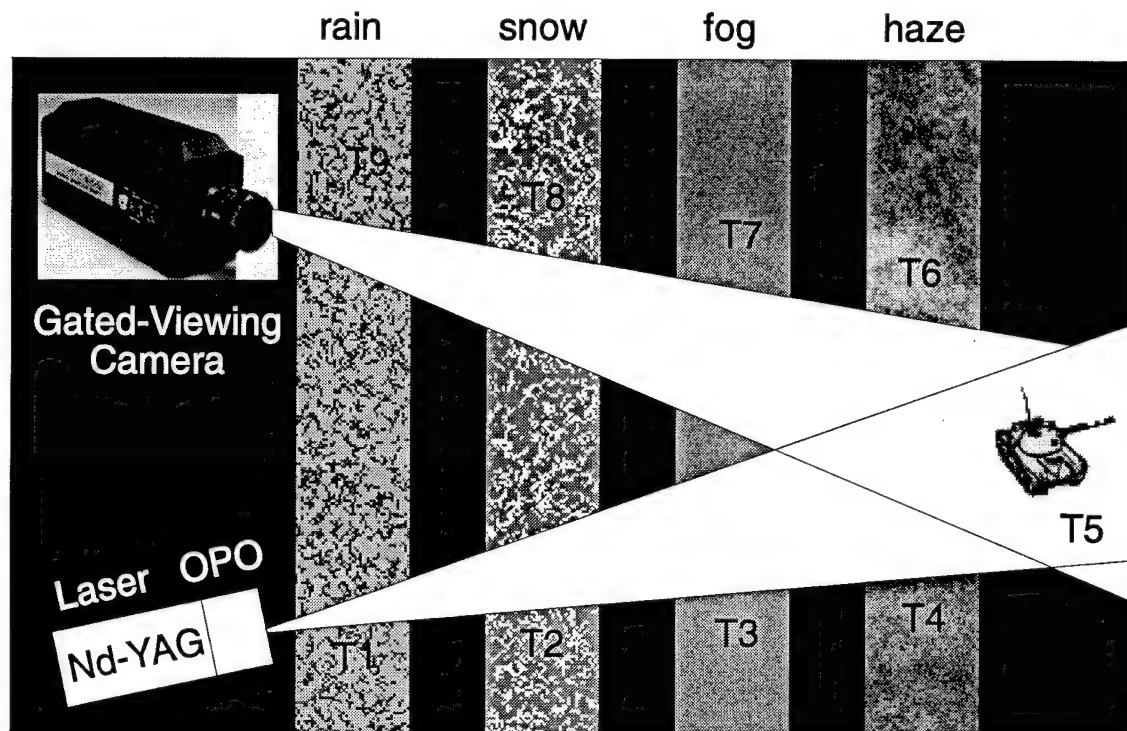
Figure 16: Principle of laser range gating

laser type	frequency-doubled Nd-YAG	Nd-YAG	Nd-YAG OPO
wavelength [nm]	532	1064	1574
pulsewidth Δt[ns]	9	9	9
pulse frequency [Hz]	30	30	30
pulses N in 100s	3000	3000	3000
CLASS 1 LIMITS			
$AEL_{singlepulse}$[J]	$2 * 10^{-7}$	$2 * 10^{-6}$	$8 * 10^{-3}$
$AEL_{multipulse}$[J]	$2.7 * 10^{-8}$	$2.7 * 10^{-7}$	$1.1 * 10^{-3}$
laser peak power [W]	≤ 3	≤ 30	≤ 120000

Table 2: Comparison of the class 1 AEL (Accessible Emission Limit) for different Nd-YAG lasers
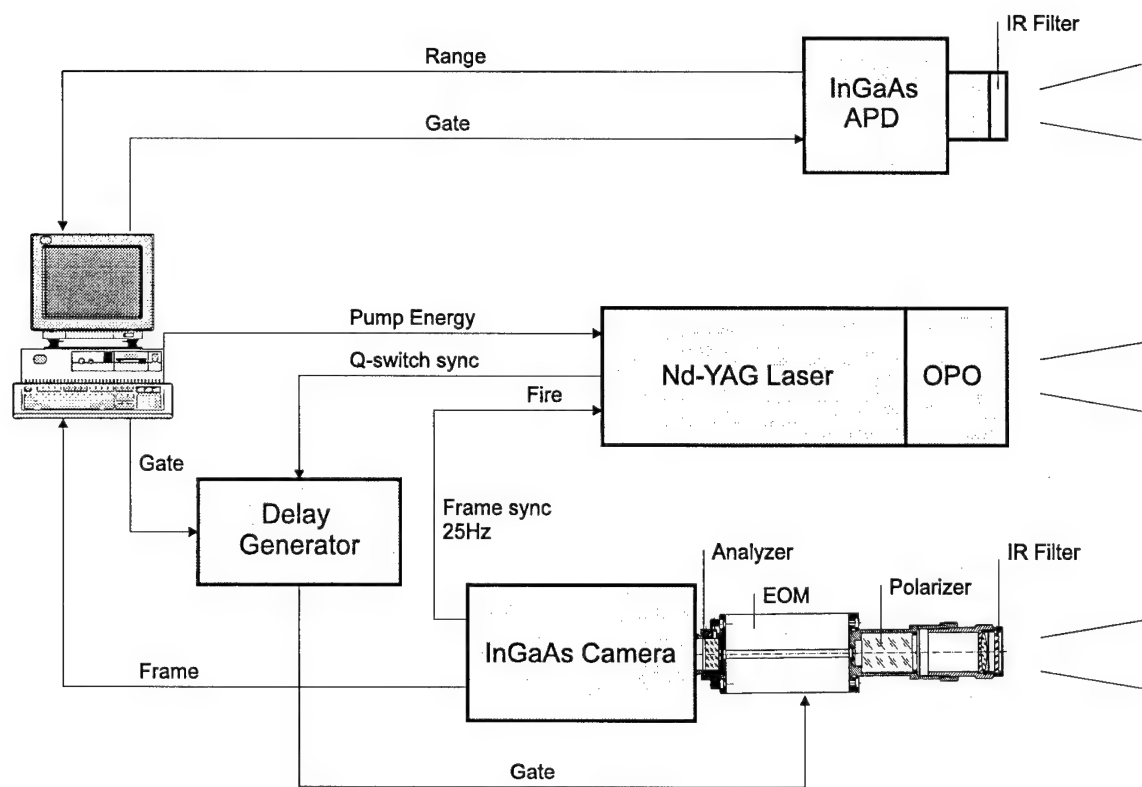
Figure 17: Design of an eye-safe laser radar

image number (Figure 19)	range [km]	gate delay [ns]	gate duration [ns]	pulse energy [mJ]	maximum grey value
(a)	0.43	2721	993	90	70
(b)	0.73	4705	993	90	224
(c)	0.88	5697	993	90	47
(d)	1.02	6689	993	90	52
(e)	1.17	7681	993	90	255
(f)	1.32	8673	993	90	97

Table 3: System parameters for the image sequence of Figure 19
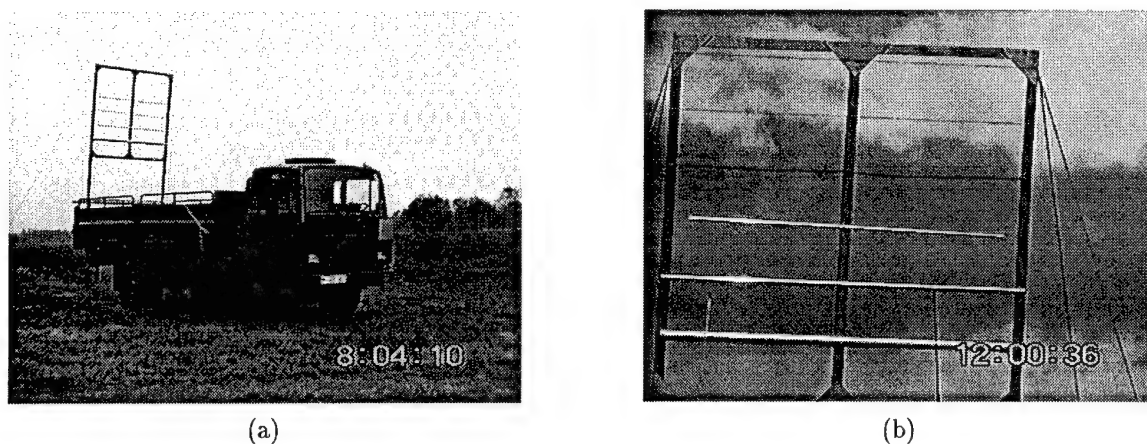
Figure 18: Video image of the truck target (a) and the special wire rack (b)
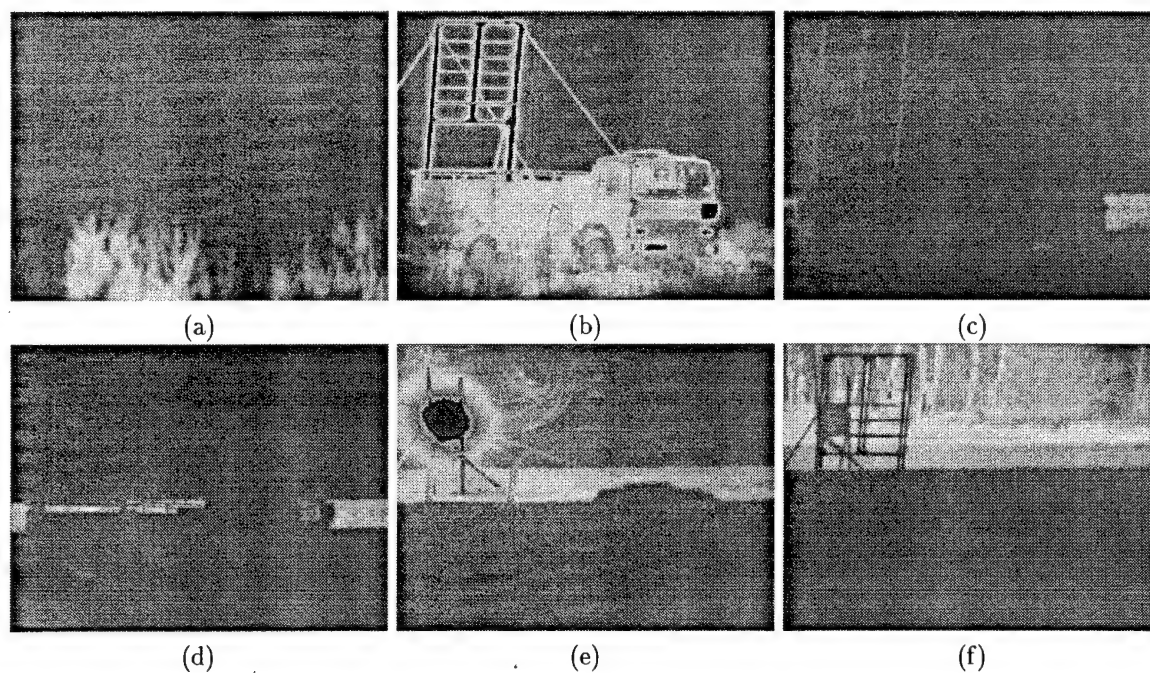
Figure 19: Sequence of range gated images at different ranges between 0.43km (a) to 1.32km (f)
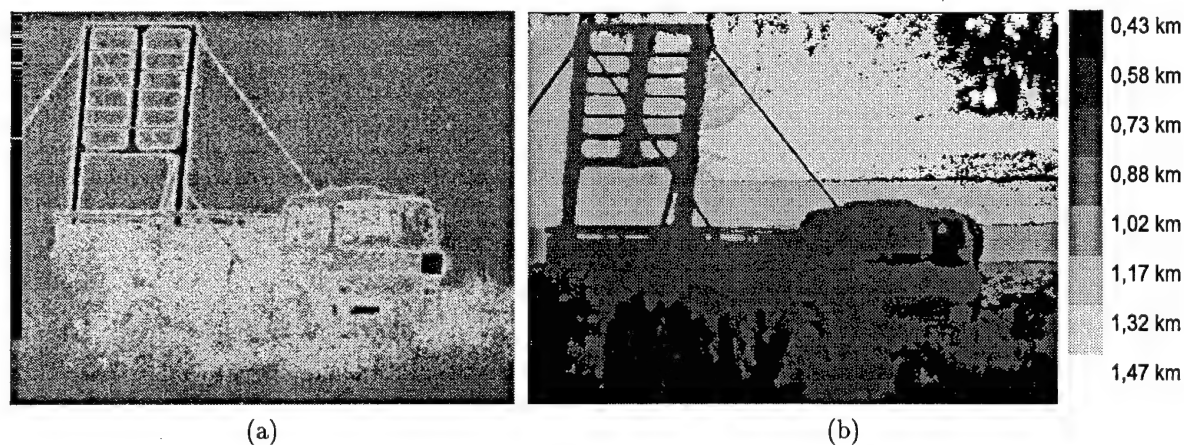

Figure 20: (a) Single range gated image (false color), high gain of 18000 and therefore a lot of excess noise; (b) Range image calculated with a threshold of 20

4.5.2 Preprocessing and Segmentation

Before executing the segmentation step it is always recommendable to concentrate on preprocessing. For laser radar image sequences there are a lot of different possiblities to enforce the quality of the single images of the sequence. Laser radar systems which are working with a MCP often generate a lot of excess noise as it can be seen in Figure 20. The excess noise depends on the MCP gain which depends mainly on the weather, the choosen range and the maximal possible laser output power. The system parameters imply the necessity of a special gain and therefore we have to live with the noise.

In the literature a lot of different noise removal filters are mentioned, which concentrate on one single image. In our case there is a much easier and more effective way. Because the process of image capturing can be controlled, we can take more images of the same range slice. Now there are three ways to combine the images:

- geometric averaging,

- median averaging or

- robust averaging.

In the example we use the geometric averaging of 20 single images for every range slice. The robust averaging would be more effective if the aim is to minimize the number of images or if there are outliers in the rows or in a whole image.

Normally the segmentation of objects is a very simple step in range gating sequences, namely a simple thresholding. The threshold only depends on the system noise and the noise removal filter or the image averaging process. Since the treshold is known, every pixel with a grey value above the threshold is an object pixel. The threshold value in the example is fixed to 22.

The correct choice of a range slice depth ΔR (30) in combination with the depth of the target ΔTR to detect is important for minimizing foreground and background pixels which are in connection with the pixels of the target. If the possible distance of the target is unknown and the range slice depth is nearly the same as the target depth, the target could be splitted in two different image slices. To be sure that one image slice of the range gated sequence includes the whole target, the range slice depth ΔR has to be two times the target depth ΔTR and the range slices have

to overlap. Therefore the exposure time Δt_{exp} is calculated to

$$\Delta t_{exp} = \frac{4 * \Delta TR}{c} - \Delta t_{puls} \qquad (32)$$

and the exposure delay step Δt_{del} has to be fixed to

$$\Delta t_{del} = \frac{2 * \Delta TR}{c} - \Delta t_{puls}. \qquad (33)$$

The combination of the laser radar system control and the digital image processing of the laser radar image sequences provides a very high effficiency of the object recognition and classification.

4.5.3 Binary Image Processing

The result of the segmentation is a binary image. The next step is the filtering of the generated binary image to delete single pixels. The single pixels, i.e. pixels with the value 1 with no other pixel in the direct neighbourhood including this value, are just set to 0 and so the number of objects stays small. The next filter is a closing filter to merge objects which are in the direct neighbourhood. It is probable that these different objects are in fact the same, because the overlapping from objects in different ranges is minimal because of the range gating.

Different objects in range gating sequences are most often connected only through the background, if the range slice depth is too large. In the example the range depth ΔR is 150m and this is very large, so we get this problem. The separation of the object from the background is only possible if the grey values of all detected objects are analysed. But in the most cases it is not necessary to do this.

The last step to separate the object is the generation of an object image, which has different numbers for every corresponding area. Two objects are distinguished when no pixel of one object is in the direct neighbourhood of each pixel of the other object.

The object image is the basis for the calculation of the object features.

4.5.4 Object Features

The choosen object features for the example are

1. Area $A[m^2]$,

2. Major Axis Length maA [m],

3. Minor Axis Length miA[m],

4. Eccentricity (≤ 1),

5. Major Axis Orientation,

6. Solidity (≤ 1),

7. Mean of the object grey values,

8. Standard Deviation of the object grey values,

9. Skewness of the object grey values,

10. Kurtosis of the object grey values

The features are separable in three groups:

- feature 1 - 3: special laser radar form features,

- feature 4 - 6: normal form features and

- feature 7 - 10: statistic moments of the grey values of each object.

The calculation of the first three features presuppose knowledge of the laser radar system design. The real size of an object pixel

$$s_{opix} = \frac{d_{pix}}{f} * R, \qquad (34)$$

with the physical width of the camera pixel d_{pix}, the focal length f and the object range R, can be used to calculate

$$
\begin{aligned}
A &= A_{pix} * s_{opix}^2, & (35) \\
maA &= maA_{pix} * s_{opix}, & (36) \\
miA &= miA_{pix} * s_{opix}. & (37)
\end{aligned}
$$

$A_{pix}, miA_{pix}, maA_{pix}$ are the corresponding pixel values to the special laser radar form features A, miA, maA. For a gating viewing camera including a MCP with a diameter d_{MCP} of 18mm and a detector field of 768x572 pixel, the physical width of the camera pixel

$$d_{pix} = \frac{d_{MCP}}{\sqrt{768^2 + 572^2}} \qquad (38)$$

can be calculated to 18.8μm. It is supposed that the diameter of the MCP is corresponding to the diagonal of the CCD chip, because the MCP is

often coupled with the CCD chip through a fiber taper.

At this moment it is possible to integrate another image processing step to reduce the number of objects. Before the calculation of the 10 object features for every object in the sequence, an elimination of small objects with a pixel area $A_{pix} \leq A_{min}$, with the threshold area A_{min}, is useful to decrease the computation time and increase the uniformity of the classes. In our example the threshold is fixed to $A_{min} = 100$.

4.5.5 Classes and Classification

The choice of the classes depends on the problem to solve. For the example we define 4 classes:

- class 1: hotspot,

- class 2: truck,

- class 3: background,

- class 4: bush (foreground).

Working with laser radar includes always the problem of hotspots, where the laser light has a strong reflection. In the laser radar sequence here, it is a retroreflector, however it also can be a windscreen or a window of a house.

The classes 3 and 4 are choosen to show the problem of an object classification. The similarity of these two classes are high and so it can be expected that there will be errors in the classification.

The method of classification used here is a weighted distance classificator, because the training set, that means the number of observations for every class, is to small for a statistical classificator. The training set is used to calculate for every class i, $i = 1...g$ of the given c classes, the mean matrix X_{ik} and the standard deviation Matrix σ_{ik} for every feature k of the given ff features.

The object feature matrix Y_{jk} for the objects j includes all object features of the o objects of the whole laser radar image sequence.

The weighted distances

$$d_{ij} = \sum_{k=1}^{ff} \sigma_{ik} * \sum_{k=1}^{ff} \frac{1}{\sigma_{ik}} (X_{ik} - Y_{jk})^2 \qquad (39)$$

can be used to find the class m of an object j while determining the minimum distance

$$dmin_j = \min(d_{ij}) \qquad \forall \quad i = 1...c. \qquad (40)$$

The object j is consequently classified to class $dmin_j = m$, $m \in 1...c$.
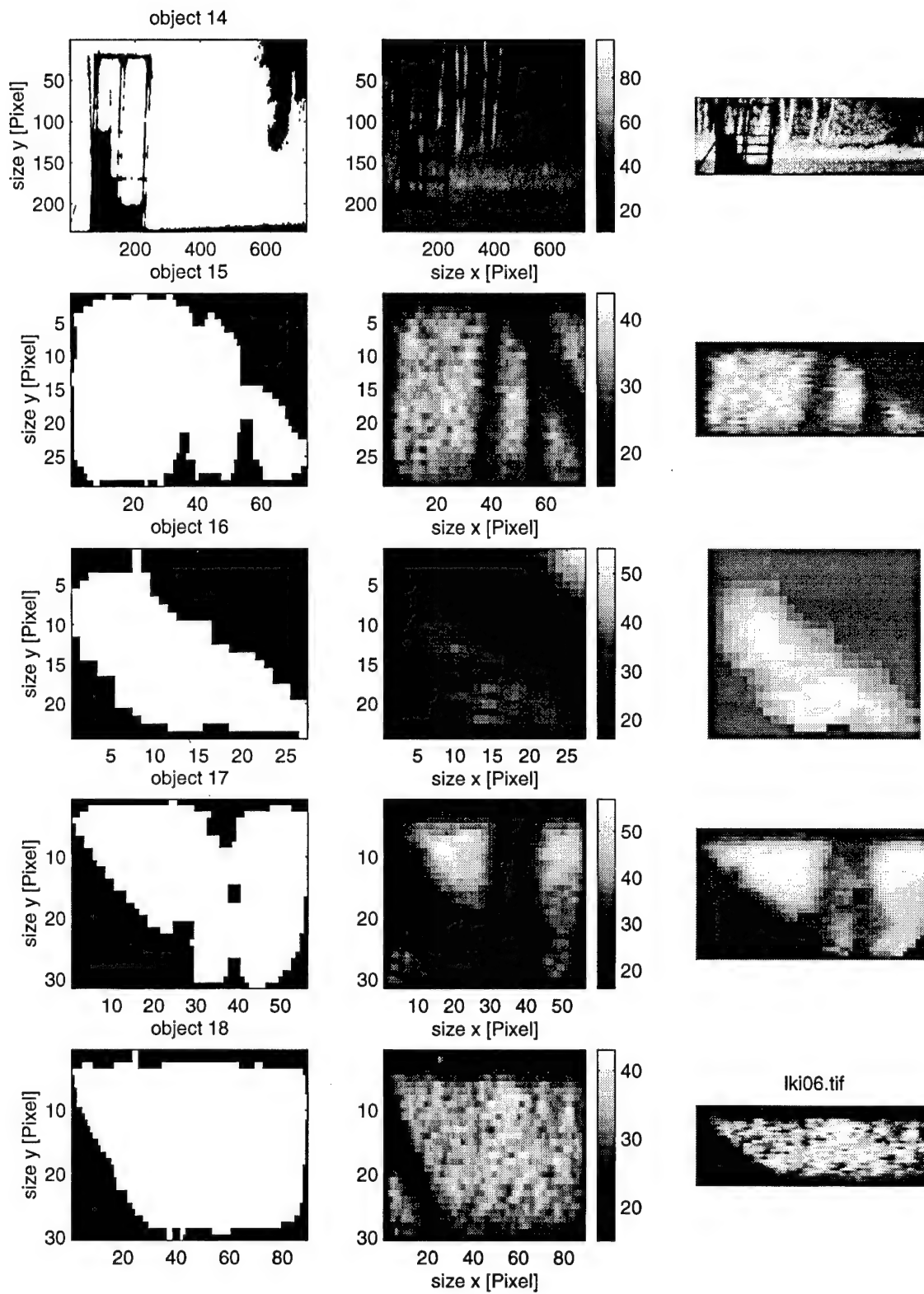
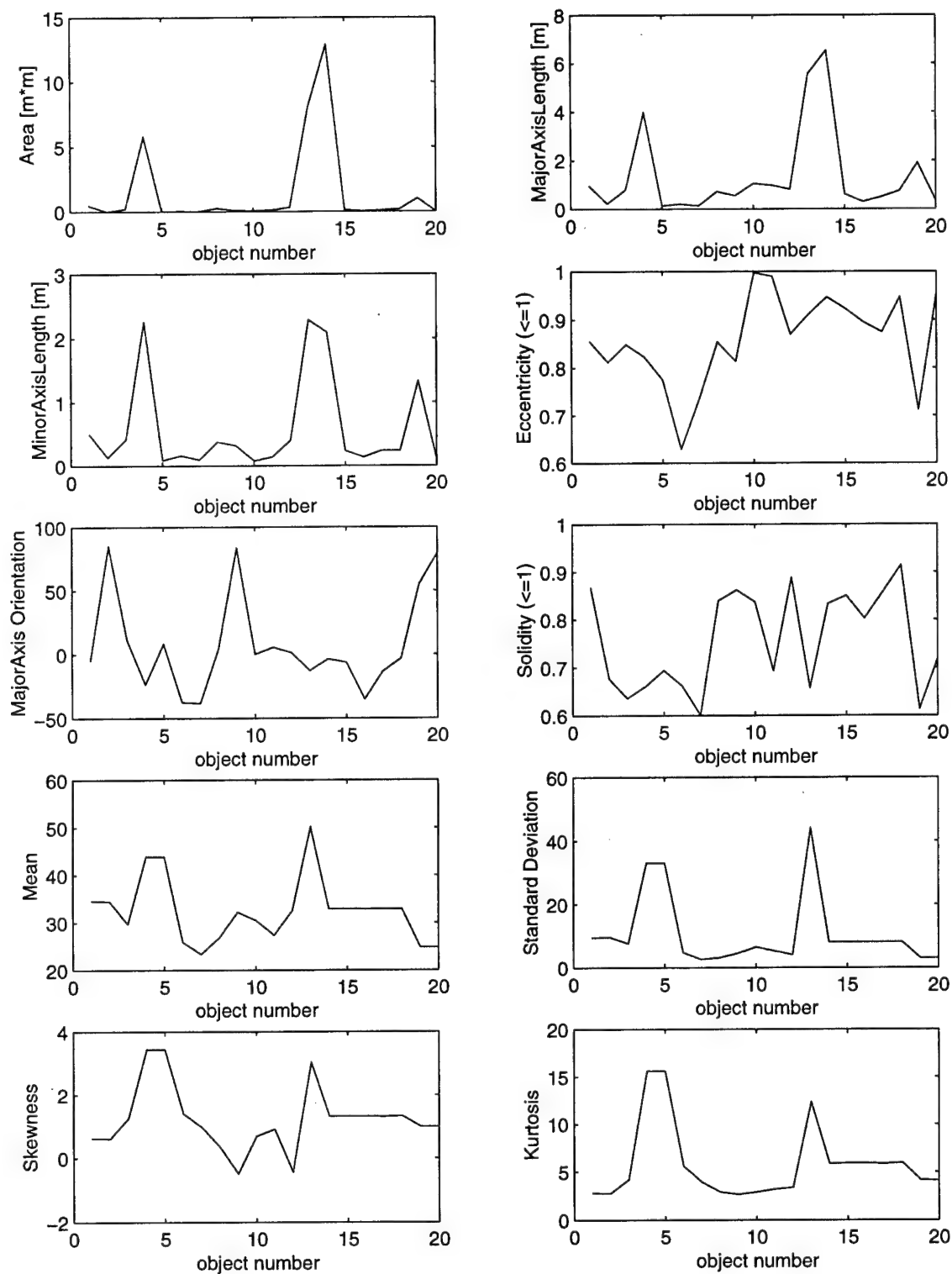Figure 21: Object 14 to 18 of the laser radar sequence in detail

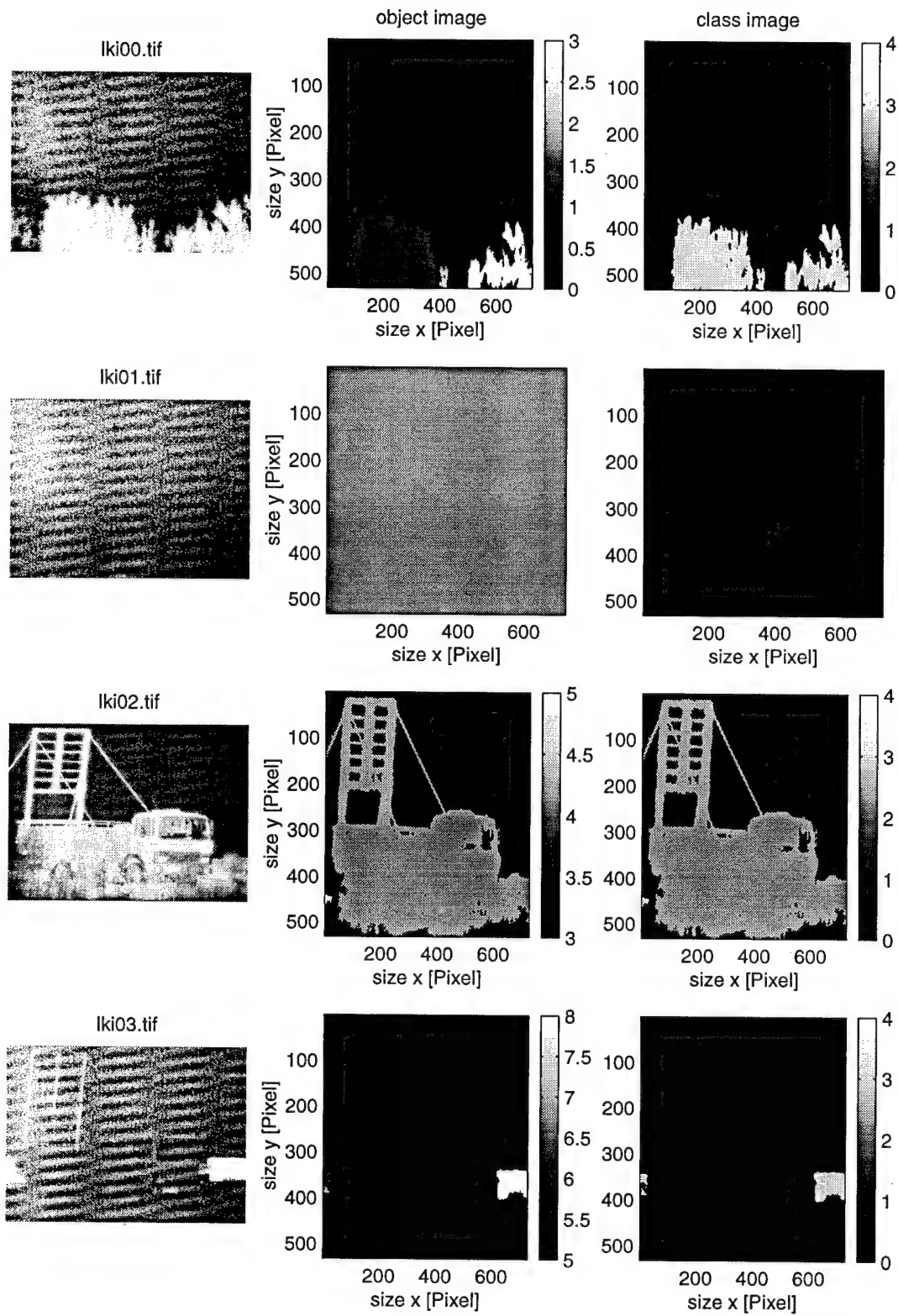Figure 22: Image features of all 20 objects of the laser radar sequence

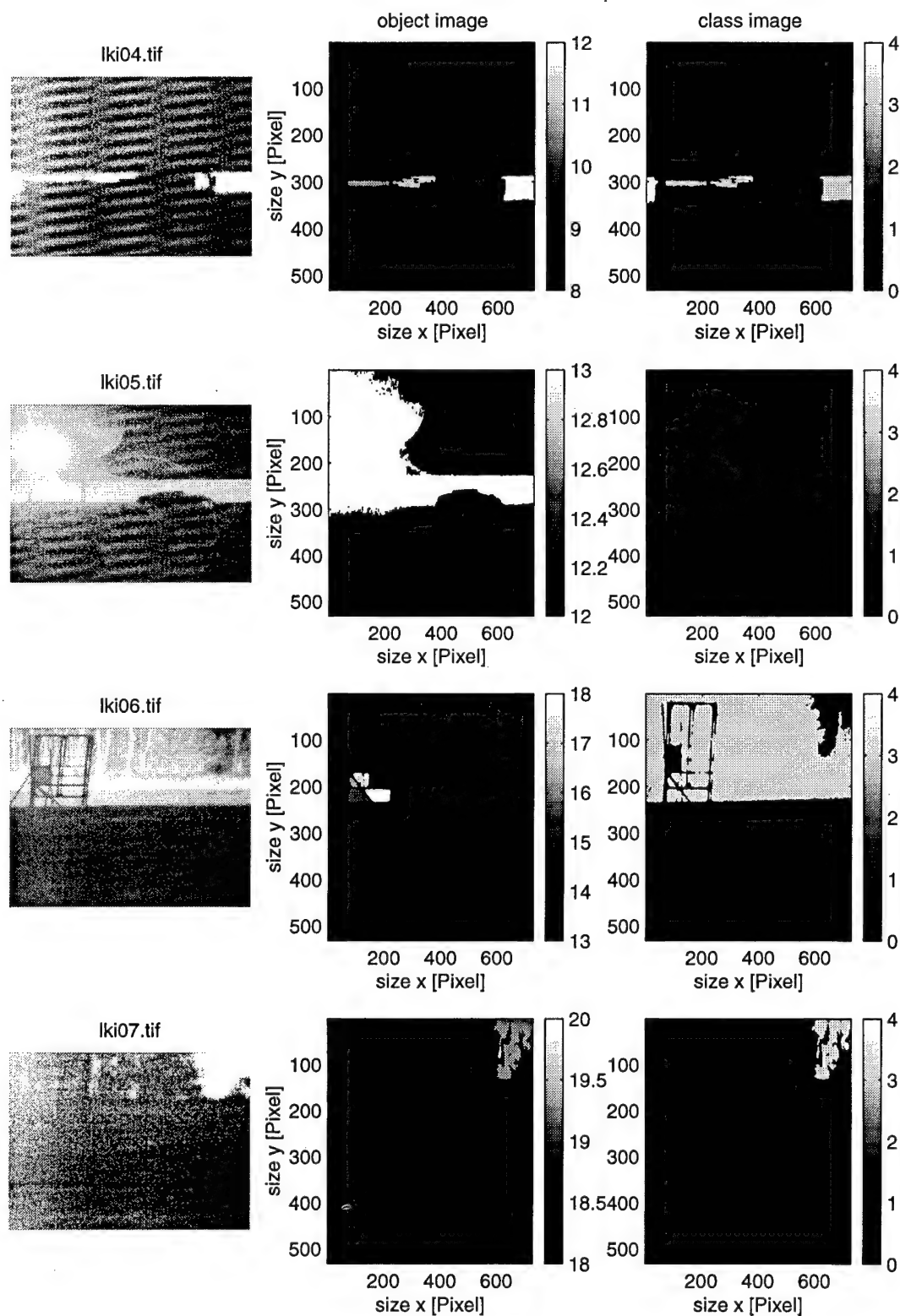Figure 23: Corresponding orginal, object and class images of the laser radar sequence, part I

object image

class image

lki04.tif

lki05.tif

lki06.tif

lki07.tif

Figure 24: Corresponding orginal, object and class images of the laser radar sequence, part II

4.5.6 Results

The rows in Figure 21 show 5 objects out of the whole laser radar image sequence, which should demonstrate the digital image processing step of the object segmentation. The first column in Figure 21 shows the binary images of object 14...18 of the laser radar image `lki06.tif` after thresholding and filtering with the closing filter. The single binary images are printed in quadratic form, but the real pixel size can be found at the axis.

The second column in Figure 21 shows the grey values of the object areas with the range of the grey values fixed in the bar on the left side of the image and the third column only the grey values of the objects. Pixels which do not belong to the object in column three are set to zero. These images are in their size real parts of the orginal laser radar image `lki06.tif`.

After the segmentation and the binary image processing 20 objects in the whole sequence are separated. The 10 diagrams of Figure 22 show the 10 object features for all 20 objects. The object number at the x axis is corresponding to the number of the objects 14...18 of Figure 21. In the most diagrams of Figure 22 there are 2 peaks for object 4 and object 13. These objects are suspicious and it should be possible to classify them easily. In fact object 4 is the truck and object 13 is the hotspot.

All images of the sequence, all objects and the whole classification result can be seen in Figure 23 and Figure 24. The first column shows the range gated images of the whole sequence. Each of the orginal images is contrast enforced to see every detail. That means that also the noise seems to be stronger if the grey value region of the individual image is small. The second and third column of Figure 23 and Figure 24 show the object and the class image to each orginal image in the same row. The bar next to the object images includes the object number and the bar next to the class images the class of each object. The lowest values in the bars mark the value for the background.

The classification result is relatively good, there are only 2 faults. The bushes in the first image of the sequence are difficult to classify, so that objects 1 and 3, which are in reality bushes are marked as background.

The classification of the truck and the hotspot is no problem.

5 Real- Time Considerations

5.1 List of symbols

Symbol	Description
C#	number of stored constants
f	ideal classificator ·
FADD#	number of additions
FDIV#	number of divisions
FMUL#	number of multiplications
k	number of groups
m_d	number of discriminant functions
m_i	number of neurons in hidden layer
m_r	number of rule in fuzzy model
n	dimension of input vectors
x	inputvector
y	outputvector

5.2 Basic Considerations

In this section the usefullness of pattern recognition methods in real time applications will be discussed. The methods discussed in "Approaches to pattern recognitions" section 4. as they are

1. discriminant analysis

2. feedforward neural network (one hidden layer)

3. first order Sugeno fuzzy model

will be examined. Based on the section 4 of "Approaches to pattern recognitions" a black box function f is defined as

$$y = f(x) \in \{1, 2, ..., k\}.$$

Function f is the ideal classification operator which separates given input vectors x into groups labeled y. It will be assumed that the applications of the three discussed pattern recognition methods approximate function f for the same process. The comparison of the three pattern recognition methods will be done regarding the following two quantities:

- number of floating point operations

- number of stored constants.

The number of floating point operations is a criterion to compare the run time of numerical algorithms independently of computation platforms. The comparison will be splitted in floating point additions (FADD), floating point multiplications (FMUL), and additionally floating point divisions (FDIV), because the run time of an FADD,

FMUL and FDIV is depending on the used hardware processing system for instance ([24],[25]). A sufficiently good approximation is counting the FADD and FMUL operations, all other floating point operations will be neglected. In addition to the number of floating point operations the memory usage (number C of stored constants) is an indicator of necessary number of memory accesses. Be k the number of groups, n the dimension of the input vector x, m_d the number of discriminant functions, m_i the number of neurons in the hidden layer of a feedforward network, m_r the number of used rules in a first order Sugeno fuzzy model. Then the number of FADD, FMUL, FDIV and stored constants C for the three different pattern recognitions method are given by:

method	FMUL#
discriminant analysis	$m_d(k+n)$
feedforward neural network	$m_i(n+k)$
first order *Sugeno* fuzzy model	$m_r(n+kn+k)$ $(+\ m_r k$ FDIV#$)$

method	FADD#
discriminant analysis	$m_d(k+n-1)$
feedforward neural network	$m_i(n-1)+n$ $+k(m_i-1)+m_i$
first order *Sugeno* fuzzy model	$km_r(n-1)+kn$ $+(m_r-1)k$

method	C#
discriminant analysis	$m_d(k+n)$
feedforward neural network	$m_i n + n + km_i + m_i$
first order *Sugeno* fuzzy model	$2m_r n + km_r n$

In the case of examples in "Approaches to pattern recognitions" section 4 the constants were defined as $n=7, k=4, m_d=2, m_i=7$, and $m_r=5$. The table below shows results of FMUL#,FADD#, and C# for a single classification.

method	FMUL#	FADD#	C#
discriminant analysis	22	20	22
feedforward neural network	77	79	91
first order *Sugeno* fuzzy model	195 (+20)	164	210

Compared to the first order *Sugeno* fuzzy model the discriminant analysis needs only about ten percent and the feedforward neural network less than fifty percent of run time. This result is convincing so that we prefer the discriminant analysis to perform single classifications.

5.3 Practical Considerations

For practical considerations the computation time of different classification algorithms will be compared. The reference computation system was a *Pentium Pro* 200 MHz *WindowsNT* PC workstation with 256 MB RAM. The classification algorithms were built using the $MATLAB^{TM}$ computation system, and the $MATLAB^{TM}Compiler$ version 5.2 .

Table 4 shows the computation time of 20 classifications using the example of section 4 in "Approaches to pattern recognitions".

method	run time (ms)
discriminant analysis	0.8
feedforward neural network	2.5
first order *Sugeno* fuzzy model	3.9

Table 4: Run time

Compared to the first order *Sugeno* fuzzy model the discriminant analysis needs only twenty percent and the feedforward neural network less than sixty six percent of run time.

Additional a PCI board with four neurochips was used to perform the classification task. This PCI board achives a peak performance of 800 MCPS (mega connections per second). Running feedforward neural network based classifications on the neurochip hardware gives a computation time of $0.28ms$ per 20 classifications.

6 Conclusions

Practical applications of pattern recognition are possible with very different approaches and the use of numerous varying feature variables. However, one has to pay attention to the fact that the hardware of the pattern recognition system is designed to meet the needs of the classification approaches.

Simple statistical descriptors and more or less complicated statistical classifiers should be chosen for first experiments. Often it turns out that they provide as well superior overall performance with respect to correctness of classification as minimal computational power requirements (time, storage capacity).

References

[1] Juergen Wernstedt, Experimentelle Prozeszanalyse, R. Oldenburg Verlag, 1989.

[2] S.L. Marple Jr., Digital spectral analysis with applications, Prentice Hall, 1987.

[3] Bernd Jaehne, Digitale Bildverarbeitung, Springer Verlag Berlin, 4. Auflage, 1997.

[4] J. C. Stover. *Optical Scattering: Measurement and Analysis.* SPIE, PO Box 10, Bellingham, Washington, USA, 2nd edition, 1995.

[5] N. I. Fisher. *Statistical Analysis of Circular Data.* University Press, Cambridge, UK, 1993.

[6] B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[7] Y. D. Willems E. P. Lafortune. Using the modified phong reflectance model for physically based rendering. Technical report, Department of Computing Science, K. U. Leuven, November 1994.

[8] J. Bienlein H. Spitzer G. Meister, R. Wiemker. In situ measurements of selected surface materials to improve analysis of remotly sensed multispectral imagery. *Proceedings of the XVIII. Congress of the International Society for Photogrammetry and Remote Sensing ISPRS*, XXXI (B7):493–498, 1996.

[9] DeBoor. *A Practical Guide to Splines.* Carl Springer, 1978.

[10] H. L. Richard, "Low cost, aircraft collision-avoidance system," in *Applied Laser Radar Technology*, G. Kamerman and W. E. Keicher, eds., *Proc. SPIE* **1936**, pp. 31–43, 1993.

[11] K. Dinndorf and D. Hayden, "Compact multichannel receiver using InGaAs APDs for single-pulse eye-safe laser radar imagery," in *Laser Radar Technology and Applications II*, G. W. Kamerman, ed., *Proc. SPIE* **3065**, pp. 22–29, 1997.

[12] T. Steiner, "Compact, 625-channel, scannerless imaging laser radar receiver," in *Laser Radar Technology and Applications I*, G. W. Kamerman, ed., *Proc. SPIE* **2748**, pp. 39–46, 1996.

[13] A. von der Fecht and H. Rothe, "Imaging laser radar experiments for enhanced- and synthetic-vision system design," in *Laser Radar Technology and Applications II*, G. W. Kamerman, ed., *Proc. SPIE* **3065**, pp. 267–278, 1997.

[14] A. von der Fecht and H. Rothe, "Comparison of imaging laser radars based on range gating under different weather conditions," in *Laser Radar Technology and Applications III*, G. W. Kamerman, ed., *Proc. SPIE* **3380**, 1998.

[15] L. W. Wolfe and G. J. Zissis, *The Infrared Handbook*, Infrared Information Analysis Center, Michigan, 1989.

[16] M. Schlessinger, *Infrared Technology Fundamentals*, Marcel Dekker Inc., New York, 1995.

[17] A. V. Jelalian, *Laser Radar Systems*, Artech House, Norwood, MA 02062, 685 Canton Street, 1992.

[18] C. S. Fox, *The Infrared and Electro-Optical Systems Handbook, Vol. 6, Active Electro-Optical Systems*, Infrared Information Analysis Center, Michigan, 1993.

[19] DIN EN 60825-1(IEC 825-1) VDE 0837, *Sicherheit von Laser-Einrichtungen*, Beuth Verlag GmbH, Berlin, 1994.

[20] K. R. Costello, "Transferred electron photocathode with greater than 20% quantum efficiency beyond 1 micron," in *Photodetectors and Power Meters II*, K. Muray and J. Kaufmann, eds., *Proc. SPIE* **2550**, pp. 177–188, 1995.

[21] D. Ricks and H. Willhite, "Handheld imaging laser radar," in *Laser Radar Technology and Applications II*, G. W. Kamerman, ed., *Proc. SPIE* **3065**, pp. 30–41, 1997.

[22] M. J. Cohen, "Room temperature InGaAs camera for NIR imaging," in *Infrared Detectors and Instrumentation*, A. M. Fowler, ed., *Proc. SPIE* **1946**, pp. 436–443, 1993.

[23] F. G. Smith, *The Infrared and Electro-Optical Systems Handbook, Vol. 2, Atmosheric Propagation of Radiation*, Infrared Information Analysis Center, Michigan, 1993.

[24] Intel, Intel Architecture Software Developer's Manual 1-3, 1997.

[25] AMD, AMD Processor datasheets, 1998.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO EN-2 AC/323 (SET) TP/1	ISBN 92-837-1001-0	UNCLASSIFIED/ UNLIMITED

5. Originator	Research and Technology Organization North Atlantic Treaty Organization 7 rue Ancelle, 92200 Neuilly-sur-Seine, France

6. Title
Advanced Pattern Recognition Techniques

7. Presented at/sponsored by
The material in this publication was assembled to support a Lecture Series under the sponsorship of the Sensors and Electronics Technology Panel and the Consultant and Exchange Programme of RTO presented on 14-15 September 1998 in Bristol, UK, on 17-18 September 1998 in Rome, Italy, and on 21-22 September 1998 in Lisbon, Portugal.

8. Author(s)/Editor(s)	9. Date
Multiple	September 1998

10. Author's/Editor's Address	11. Pages
Multiple	168

12. Distribution Statement	There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Pattern recognition	Neural nets
Target recognition	Fuzzy sets
Image restoration	Imaging
Image processing	Computation
Inversion techniques	Computerized simulation
Land mine detection	Measurement
Statistical analysis	

14. Abstract

Pattern recognition is the extraction of consistent information from noisy spatiotemporal data. It can be and is currently being used in systems for battlefield supervision, smart weapons, and anti-counterfeiting of all kinds. A current application is the automatic detection of land mines and unexploded ordance (UXO).

The methods employed can be subdivided in the following manner: (i) statistical methods, (ii) neuro-methods, (iii) fuzzy-methods, and (iv) neuro-fuzzy methods. Each of these methods has its special advantages and drawbacks, but all of them require the computation of feature variables from measurement or simulation data, e.g. from microwave backscattering.

The Lecture Series covers the following topics:
- Introductory Overview on Pattern Recognition Techniques, (i)-(iv)
- Feature Extraction for Pattern Recognition by
  - Electromagnetic, magnetic, and acoustic singularity identification
  - Model based scattering signatures
  - Wavelet techniques
  - SAR/ISAR imaging
  - Bistatic microwave imaging
  - Electromagnetic inversion techniques
- Real-time Implementation of Pattern Recognition Methods
- Introduction to Software and Hardware for Pattern Recognition

This Lecture Series, sponsored by the Sensors and Electronics Technology Panel (SET) of RTO, has been implemented by the Consultant and Exchange Programme.

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie 0(1)55.61.22.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

RTO NON CLASSIFIEES

L'Organisation pour la recherche et la technologie de l'OTAN (RTO), détient un stock limité de certaines de ses publications récentes, ainsi que de celles de l'ancien AGARD (Groupe consultatif pour la recherche et les réalisations aérospatiales de l'OTAN). Celles-ci pourront éventuellement être obtenues sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous par lettre ou par télécopie à l'adresse indiquée ci-dessus. Veuillez ne pas téléphoner.

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de la RTO et de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Directeur - Gestion de l'information
  (Recherche et développement) - DRDGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
RTO Coordination Office
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076
Etats-Unis

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Les demandes de documents RTO ou AGARD doivent comporter la dénomination "RTO" ou "AGARD" selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
    http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE
FRANCE

Telefax 0(1)55.61.22.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED
RTO PUBLICATIONS

NATO's Research and Technology Organization (RTO) holds limited quantities of some of its recent publications and those of the former AGARD (Advisory Group for Aerospace Research & Development of NATO), and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. **Please do not telephone**.

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO publications, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your organisation) in their distribution.

RTO and AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Director Research & Development
 Information Management - DRDIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
RTO Coordination Office
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

SALES AGENCIES

NASA Center for AeroSpace
 Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076
United States

The British Library Document
 Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Canada Institute for Scientific and
 Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
   http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic
Database or on CD-ROM)

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7